

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

NIVEAU D'ABSTRACTION
DANS L'EXPLORATION DE DONNÉES MUSICALES

THÈSE
PRÉSENTÉE COMME EXIGENCE PARTIELLE
DU DOCTORAT EN INFORMATIQUE COGNITIVE

PAR
LOUIS ROMPRÉ

FÉVRIER 2016

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.07-2011). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Dans un premier temps, je tiens à remercier Ismaïl Biskri et Jean-Guy Meunier qui ont accepté de m'encadrer en tant que directeur et codirecteur de recherche. Vous m'avez enseigné les fondements de ce qui a mené à l'écriture de cette thèse. Vos conseils et votre appui m'ont toujours aidé à progresser dans l'accomplissement de ce projet. Merci à mon fils Alex et ma fille Mélia, vous avez été pour moi une source d'inspiration et de motivation. Merci également à ma conjointe Émilie qui m'a offert son soutien et qui m'a accompagné durant toutes les années qu'a duré ce travail. Je tiens également à souligner la contribution de l'UQAM et à son partenaire Hydro-Québec qui ont souligné mes efforts en m'octroyant trois bourses d'excellence. Finalement, je tiens à remercier mes parents, Louis-Jules et Thérèse qui m'ont toujours appuyé et encouragé dans mes démarches. À vous tous, merci.

TABLES DES MATIÈRES

LISTE DES FIGURES	ix
LISTE DES TABLEAUX	xiii
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	xv
RÉSUMÉ.....	xvii
CHAPITRE I	
PROBLÉMATIQUE DE RECHERCHE	1
1.1. Revue de la littérature	2
1.1.1. L'annotation	2
1.1.2. L'extraction de caractéristiques et la classification	6
1.1.3. Le filtrage collaboratif et la recommandation	7
1.1.4. Le fingerprinting	8
1.2. État de la situation actuelle	10
1.3. Solution proposée	11
1.4. Contribution	14
1.5. Structure du document	15
CHAPITRE II	
LA MUSIQUE ET SES FORMES DE REPRÉSENTATION	19
2.1. Les descripteurs.....	19
2.1.1. Les descripteurs de bas niveau.....	21
2.1.2. Les descripteurs intermédiaires	26
2.1.3. Les descripteurs de haut niveau	31
2.2. Le traitement du signal	33
2.2.1. Le domaine spatial.....	35
2.2.2. Le domaine spectral.....	38
2.2.3. Les bibliothèques dédiées à l'extraction des caractéristiques.....	44
2.3. L'acquisition de données subjectives	48

CHAPITRE III	
LA CLASSIFICATION	49
3.1. La préparation des données	51
3.1.1 Les données descriptives	57
3.2. L'application d'une méthode de classification	58
3.2.1. La classification selon le voisinage	59
3.2.2. Le clustering.....	61
3.2.3. Les réseaux de neurones	64
3.2.4. Les SVM	73
3.2.5. Discussion au sujet du MIR et de la classification.....	76
3.3. L'interprétation des résultats	80
CHAPITRE IV	
LES RÈGLES D'ASSOCIATION	83
4.1. Définition des règles d'association	85
4.1.1. Itemset.....	86
4.1.2. Support d'un itemset.....	88
4.1.3. Les règles d'association	91
4.1.4. Le support d'une règle d'association	93
4.1.5. La confiance d'une règle d'association.....	94
4.2. L'extraction des règles d'association	95
4.3. Les règles d'association et la musique.....	105
CHAPITRE V	
SOLUTION PROPOSÉE AU PROBLÈME D'EXPLORATION	109
5.1. Solution proposée.....	110
5.2. Considérations architecturales	112
5.3. Architecture proposée	115
5.4. Spécification du prototype développé.....	118
CHAPITRE VI	
EXPÉRIMENTATIONS ET DISCUSSION	121
6.1. Expérimentation I : évaluation de classifications existantes	122
6.1.1. Méthodologie	123

6.1.2.	Ensemble de données.....	124
6.1.3.	Résultats et discussion.....	125
6.2.	Expérimentation II : classification à l'aide de règles d'association	130
6.2.1.	Méthodologie.....	131
6.2.2.	Ensemble de données.....	138
6.2.3.	Résultats et discussion.....	139
6.3.	Expérimentation III : analyse de la structure de pièces musicales	152
6.3.1.	Méthodologie.....	153
6.3.2.	Ensemble de données.....	155
6.3.3.	Résultats et discussion.....	155
6.4.	Expérimentation IV : recommandation d'artistes	169
6.4.1.	Méthodologie.....	170
6.4.2.	Ensemble de données.....	173
6.4.3.	Résultats et discussion.....	174
6.5.	Discussion.....	178
CHAPITRE VII		
CONCLUSION.....		183
7.1.	Perspectives.....	185
ANNEXE A		
EXEMPLE DE REQUÊTE SQL EXPLOITANT LES RÈGLES D'ASSOCIATION		187
BIBLIOGRAPHIE.....		193

LISTE DES FIGURES

Figure	Page
1.1. Nuage de tag représentant des genres musicaux	3
1.2. Nuage de tag représentant des artistes	5
1.3. Spectrogramme d'une pièce musicale.....	9
1.4. Empreinte fréquentielle d'une pièce musicale	9
2.1. Niveaux d'abstraction des descripteurs utilisés pour décrire la musique.....	20
2.2. Illustration d'une onde sonore pure	22
2.3. Propriétés de l'onde sonore	23
2.4. Onde sonore complexe.....	24
2.5. Onde sonore de la note E (mi) jouée à la guitare	25
2.6. Onde sonore de la note E (mi) jouée à la basse	25
2.7. Données textuelles périphériques aux données sonores dans iTunes.....	30
2.8. Forme de l'information du domaine spatiale	34
2.9. Forme de l'information du domaine spectrale	34
2.10. Échantillonnage de l'onde sonore	35
2.11. Distance entre deux attaques.....	37
2.12. Décomposition d'un signal en terme de sinusoïdes	38
2.13. Courbe de correspondance Mel – Fréquence	42
2.14. Filtre triangulaire à l'échelle de Mel.....	43
2.15. Les schémas de classification diffèrent d'un individu à l'autre	47
3.1. Processus de classification numérique.....	50
3.2. Création des représentations vectorielles.....	56
3.3. Classification basée sur le plus proche voisin	59
3.4. Classification basée sur les k plus proches voisins avec $k = 3$	60

3.5. Schéma d'un neurone artificiel.....	64
3.6. Exemple d'un réseau de neurones.....	67
3.7. Processus de classification supervisée.....	69
3.8. Les SVM	74
3.9. Transposition dans une autre dimension	76
4.1. Diagramme de Hasse.....	87
4.2. Élagage de l'itemset non fréquent CB et de ses descendants	100
4.3. Itemsets fréquents déduits de {Am, C, G}.....	101
4.4. Nettoyage des règles d'association.....	104
5.1. Architecture modulaire.....	114
5.2. Architecture proposée	116
6.1. Séquence des opérations de l'expérimentation I	123
6.2. Répartition des pièces présumées BLUES	126
6.3. Répartition des pièces présumées CLASSIQUE.....	126
6.4. Répartition des pièces présumées COUNTRY	126
6.5. Répartition des pièces présumées DISCO	126
6.6. Répartition des pièces présumées HIP-HOP.....	127
6.7. Répartition des pièces présumées JAZZ	127
6.8. Répartition des pièces présumées MÉTAL.....	127
6.9. Répartition des pièces présumées POP	127
6.10. Répartition des pièces présumées REGGAE	128
6.11. Répartition des pièces présumées ROCK.....	128
6.12. Processus de la création de la représentation vectorielle des fichiers	132
6.13. FFT appliquée à un signal périodique.....	133
6.14. FFT appliquée à un signal non périodique.....	134
6.15. FFT appliquée à un signal filtré.....	135
6.16. Élimination des bandes de fréquences jugées non significatives	137
6.17. Capture d'écran du module d'extraction des descripteurs de bas niveau	138

6.18. Schéma utilisé comme repère lors de l'expérimentation II	139
6.19. Processus d'analyse de la structure d'une pièce musicale	154
6.20. Forme de l'onde sonore de la pièce Adrénaline.....	156
6.21. Découpage de l'onde sonore de la pièce Adrénaline.....	157
6.22. Graphique illustrant la structure de la pièce Adrénaline.....	158
6.23. Découpage de la structure de la pièce Adrénaline.....	159
6.24. Graphique A de la pièce « Haircuts uniform »	161
6.25. Graphique B de la pièce « Haircuts uniform »	162
6.26. Superposition des graphiques A et B. de la pièce « Haircuts uniform »	163
6.27. Graphique de la pièce « My body is a cage ».....	164
6.28. Localisation de la phrase « My body is a cage » et « I'm living in a cage »...168	
6.29. Profondeur des transactions.....	171
6.30. Capture d'écran du module d'exploration d'artistes similaires	172
6.31. Lecteur audio.....	181

LISTE DES TABLEAUX

Tableau	Page
2.1. Genres attribués par iTunes et Archambault.ca en 2012	28
2.2. Genres attribués par iTunes et Archambault.ca en 2015	29
2.3. Boîtes à outils d'extraction et propriétés couvertes	45
3.1. Répartition de caractéristiques à l'intérieur des membres d'une classe.....	52
3.2 : Calcul du degré de similarité des membres d'une catégorie.....	54
3.3. Degré d'appartenance des membres à leur catégorie.....	54
3.4. Résultat du concours MIREX 2014	79
4.1. Répartition des items dans les transactions	86
4.2. Support brut et support.....	89
4.3. Mesure d'intérêt des règles d'association.....	92
4.4. Mesure de support.....	93
4.5. Mesure de confiance.....	95
4.6. Support des sous-ensembles candidats.....	104
4.7. Confiance des règles candidates.....	104
6.1. Découpage du spectre fréquentiel en bande de fréquence	136
6.2. Règles extraites pour un sous ensemble présumé classique	140
6.3. Règles extraites pour un sous ensemble présumé rock.....	140
6.4. Pourcentage de cohérence avec GTZAN	141
6.5. Degré de similarité de 20 pièces musicales présumées CLASSIQUE.....	144
6.6. Degré de similarité de 20 pièces musicales présumées ROCK	145
6.7. Degré de similarité de 20 pièces musicales présumées HIP-HOP.....	146
6.8. Règles d'association de deux pièces présumées CLASSIQUE.....	148
6.9. Règles d'association de deux pièces présumées ROCK.....	149

6.10. Règles d'association de deux pièces présumées ROCK et MÉTAL	150
6.11. Nombre de règles d'association extraites par segment de 5 secondes.....	166
6.12. Artistes similaires à Jacques Brel selon Echonest.....	174
6.13. Artistes jugés similaires en fonction des règles d'association.	175
6.14. Règles d'association utilisées pour dégager de nouveaux artistes.....	176
6.15. Relation entre support et nombre d'artistes dégagés	177

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

DBN	<i>Deep Believe Network</i> est un modèle de réseau de neurones.
DCN	<i>Deep Convolutional Network</i> est un modèle de réseau de neurones.
ENEST	Ensemble de données musicales accessible par un appel au service web Echnonest.
FFT	Transformée de Fourier rapide (Fast Fourier Transform).
GMM	Modèle de classification. L'acronyme GMM provient de l'expression anglaise Gaussian Mixture Model.
GTZAN	Ensemble d'extraits de pièces musicales rassemblées par George Tzanetakis et utilisé dans le cadre de l'article « Musical Genre Classification of Audio Signals ».
ISMIR	<i>International Society for Music Information Retrieval</i> (ISMIR) est une conférence internationale dédiée au MIR.
LAT	La durée logarithmique de l'attaque est une mesure associée au domaine spatial. Cette mesure est utilisée dans le traitement du signal. L'acronyme LAT provient de l'expression anglaise <i>Log Attack Time</i> .
MFCC	Mesure des coefficients ceptraux. C'est une mesure du domaine spectral utilisée dans le traitement du signal. L'acronyme MFCC provient de l'expression anglaise <i>Mel Frequency Ceptral Coefficients</i> .
MIR	Est un domaine de recherche interdisciplinaire dédiée à la recherche de similarité musicale. L'acronyme MIR provient de l'expression anglaise <i>Music Information Retrieval</i> .

MIREX	<i>Music Information Retrieval Evaluation eXchange</i> (MIREX) est un concours organisé dans le cadre de la conférence ISMIR dont l'objectif est de tester de nouveaux systèmes de traitement de données musicales.
SVM	Modèle de classifieur numérique à apprentissage supervisé (Support Vector Machine)
UNIF	Unité d'information dépendante du domaine d'application qui est utilisée dans le traitement automatisé de données numériques.
ZCR	Le taux de passage par zéro est une mesure associée au domaine spatial. Cette mesure est utilisée dans le traitement du signal. L'acronyme ZCR provient de l'expression anglaise <i>Zero Crossing Rate</i> .

RÉSUMÉ

La quantité de données musicales créées et diffusées est si importante qu'il est difficile d'explorer toutes ces données. Au cours des dernières années, plusieurs stratégies ont été proposées afin de soutenir l'exploration des données musicales. Quoique plusieurs de ces stratégies donnent des résultats intéressants, elles ont tendance à se spécialiser dans le traitement d'une seule facette de l'exploration des données musicales. Elles proposent des mécanismes spécifiques que ce soit pour organiser, analyser ou accéder aux données. Par conséquent, l'exploration de données musicales nécessite l'utilisation de plusieurs outils complémentaires.

La musique peut être abordée selon différentes perspectives. Chacune de ces perspectives donne un sens particulier à la musique. De nombreuses évidences suggèrent que l'expérience musicale implique de nombreux processus cognitifs qui manipulent différentes formes d'informations. Dans le cadre de cette thèse, il est proposé de grouper ces informations en fonction de leur niveau d'abstraction. Il est présumé que l'expérience musicale dépend fortement du traitement d'information appartenant à trois différents niveaux d'abstraction. Chacun de ces niveaux contribue de manière distincte à définir l'expérience musicale. Le premier niveau d'abstraction est constitué de descripteurs de bas niveau qui réfèrent à des propriétés de l'onde sonore. Ces descripteurs renseignent sur la nature de la source sonore. Le second niveau d'abstraction est composé de descripteurs qui réfèrent aux nominations attribuées à des séquences particulières de sons et aux données périphériques liées à ces séquences. Ces descripteurs renseignent sur les conventions sociales et culturelles qui influencent la manière dont la musique est décrite. Le troisième et dernier niveau est composé de descripteurs qui expriment les préférences, les motivations et les connaissances des individus. C'est uniquement en considérant les descripteurs de ce niveau que la valeur réelle accordée par un individu à une musique peut être capturée.

Cette thèse porte sur l'usage des règles d'association pour soutenir l'exploration de données musicales. L'utilisation de règles d'association se présente comme un moyen de faciliter la mise en œuvre de différentes stratégies vouées à soutenir l'exploration de données musicales. Cette méthode de forage de données a la capacité de dégager des connaissances cachées dans des données dont la nature peut varier. Par conséquent, il devient possible de concevoir un

système d'aide à l'exploration de données musicales capable d'une part de faciliter la mise en œuvre de différentes stratégies et d'autre part de traiter des descripteurs liés à plus d'un niveau d'abstraction.

MOTS-CLÉS : Data mining, Musique, Règles d'association, Classification, Traitement de données, Exploration, MIR.

CHAPITRE I

PROBLÉMATIQUE DE RECHERCHE

La numérisation du son a favorisé le stockage de données sonores et par conséquent la création de nombreuses bases de données de documents audio. Rapidement, ces bases de données sont devenues gigantesques. Certaines d'entre elles sont composées de plusieurs millions de documents audio. Les boutiques iTunes, GooglePlay et Amazon forment quelques exemples de ces vastes bases de données. Dans son rapport annuel publié en 2015, la fédération internationale de l'industrie phonographique¹ (IFPI) a estimé que les catalogues de ces grandes boutiques de musique numérisée étaient composés de plus de 37 millions de titres. Devant une telle quantité de documents, l'exploration des données sonores peut devenir une tâche complexe qui demande un effort considérable. Par conséquent, certains titres peuvent demeurer peu ou pas explorés et parmi ces titres certains peuvent être d'un grand intérêt.

L'engouement pour les bases de données de documents audio ne cesse de croître. Signe de cet engouement, Apple annonçait en février 2013 avoir procédé à la vente de 25 milliards de pièces musicales à partir d'iTunes². La quantité de données sonores qui transigent sur les réseaux informatiques est également en pleine croissance. Les téléchargements de musique, l'utilisation de service de téléphonie IP et la diffusion en flux contribuent à ce phénomène. La popularité des tablettes électroniques et des téléphones mobiles accélère cette expansion.

¹ <http://www.ifpi.org/>

² <http://www.apple.com/pr/library/2013/>

Ces appareils sont dans l'ensemble dotés de fonctionnalités qui facilitent l'enregistrement et le partage des données sonores. Les réseaux sociaux contribuent également au phénomène, car ils encouragent la diffusion des documents sonores. Les nombreuses données audio collectées sont souvent emmagasinées de manières désorganisées. L'un des défis actuels est de proposer des mécanismes efficaces permettant d'améliorer l'exploration des données sonores.

1.1. Revue de la littérature

Au cours des dernières années, plusieurs stratégies ont été proposées afin de soutenir la recherche de similarités musicales (Tzanetakis et Cook, 2002, Wang, 2003, Pampalk et al., 2003 ; Pérez-Sancho et al., 2009; Aucouturier et Pachet, 2003). Sommairement, l'objectif de ces travaux est d'organiser les données de manière à en faciliter l'accès. Pour atteindre cet objectif, différentes méthodes ont été suggérées. L'annotation (Bertin-Mahieux et al., 2010), la classification (Mittra et Wang, 2007), le filtrage collaboratif (Linden et al., 2003) et le *fingerprinting* (Wang, 2003) représentent quelques-unes de ces méthodes. Elles permettent respectivement de favoriser la recherche, d'organiser les données, de faire des recommandations et de reconnaître des musiques.

1.1.1. L'annotation

L'une des principales méthodes suggérées est l'annotation. Cette méthode consiste à ajouter aux documents sonores des données textuelles descriptives. L'objectif est d'accélérer et de faciliter la recherche. Pour ce faire, des métadonnées sont identifiées et inscrites de manière à favoriser le rappel des

documents lors de recherches (Bertin-Mahieux et al., 2010). Les métadonnées utilisées expriment généralement des informations relatives au contenu (p. ex. le genre musical) ou à des informations contextuelles (p. ex. le titre, l'auteur, l'année de publication, etc.). Les métadonnées peuvent être présentées à l'utilisateur sous différentes formes. Par exemple, elles peuvent être présentées sous la forme d'un nuage de tags (figure 1.1 et 1.2). Dans ce contexte, la taille du texte renseigne l'utilisateur sur la popularité du tag. Plus la taille du texte est importante, plus la popularité du tag est grande.

Alternative/Nu-Metal
 Black/Folk/Pagan/Viking Blues
 Breakbeat/Dubstep
 Chillout/Lounge/Downtempo
 Classic/Hard Dance Pop/EuroDance
 Heavy/Power/Progressive House
 Indie/Post-Rock/Post-Punk Jazz
 Metalcore/Deathcore/Mathcore Pop
 Pop-Rock/Soft Rock
 Progressive/Art-Rock Punk
 Rap/Hip Hop Synthpop / New Wave
 Thrash/Speed Trance

1.1. Nuage de tag représentant des genres musicaux

La taille du texte représente la popularité du genre.

Source de l'image : www.plixid.com

L'annotation manuelle est la méthode la plus simple. Pour annoter les pièces musicales, des utilisateurs doivent préalablement écouter les différents documents audio afin d'être en mesure de définir les métadonnées qui seront les plus représentatives. Les connaissances musicales et culturelles de la personne responsable d'annoter les documents influencent fortement la nature et même la qualité des métadonnées sélectionnées (Bertin-Mahieux et al., 2010).

Même dans un scénario où les métadonnées utilisées représentent adéquatement le contenu des documents sonores, leur utilisation pour orienter une recherche au sein d'une importante collection de documents sonores demande d'une part à ce que le requérant ait une idée précise de ce qu'il recherche et d'autre part que les mots clés utilisés lors de la recherche concordent avec les métadonnées annotées aux documents recherchés (Grosche et al., 2012). De plus, dans un contexte où il y a une surabondance de données musicales, l'annotation de document ne peut pas être réalisée manuellement. L'une des solutions pour répondre à cette problématique est de considérer le contenu des documents afin d'automatiser la détection de similarités musicales.

10cc 2012 Adele Aerosmith
 Arctic Monkeys Beatles Blur
 Cat Power CELTIC Coldplay
 David Bowie Deep Purple Depeche Mode
 Frank Ocean Gandalf Gotye
 Iron Maiden Jason Mraz Joss Stone
 Justin Bieber Kiss Lana Del Rey
 Led Zeppelin LINKIN PARK Madonna
 Marillion Maroon 5
Mike Oldfield Motorhead
 Muse OASIS Pet Shop Boys Pink Floyd
 Queen Radiohead Red Hot Chili Pepp...
 Reggae Rick Ross Rihanna Santana
 Skrillex Slipknot Southern Rock
 Testament THE BEATLES The Cure
 The Killers The Vaccines The Xx
 Two Door Cinema Cl...

1.2. Nuage de tag représentant des artistes

La taille du texte représente la popularité de l'artiste.
 Source de l'image : www.plixid.com

1.1.2. L'extraction de caractéristiques et la classification

Une deuxième méthode employée pour soutenir l'exploration de données musicales est la classification automatisée. L'onde sonore est caractérisée par des fréquences, une durée, des amplitudes, etc. Il est possible d'utiliser ces caractéristiques pour représenter le contenu d'un document audio. Ainsi, un document peut être jugé similaire à un autre document ou à un ensemble de documents en fonction des caractéristiques qu'il partage avec eux. Plusieurs travaux portent sur l'extraction et l'utilisation des caractéristiques de l'onde sonore à des fins de recherche de similarités musicales (Grosche et al., 2012; Wold et al., 1996; Tzanetakis et al., 2002). Dans la littérature, la procédure qui mène à l'extraction de ces caractéristiques est sensiblement la même d'une approche à l'autre. L'onde sonore est d'abord segmentée. Ensuite, divers algorithmes sont appliqués aux segments afin de dégager certaines caractéristiques. Un vecteur de n caractéristiques est créé pour tous les segments. Une représentation vectorielle est ainsi créée pour chacun des documents. Ces représentations vectorielles sont alors utilisées comme intrants pour des classifieurs numériques.

Les réseaux de neurones artificiels sont fréquemment utilisés pour réaliser la classification (Mitra et Wang, 2007). Cet intérêt provient de leur capacité à capturer des propriétés structurantes à même les données brutes (Anderson, 1995; Haykin, 1994). Bien que l'utilisation de réseaux de neurones puisse mener à des résultats intéressants, leur opacité pose problème. Les résultats changent d'un réseau de neurones à l'autre. Une fois le réseau conditionné, il est difficile de savoir quelles sont les caractéristiques importantes et celles qui ne le sont pas. La contribution d'une unité d'information dans la définition d'une classe peut être difficile à établir. De plus, il n'existe pas de théorie formelle pour

prédire qu'elles sont les unités d'information à considérer (Estes, 1994). La plupart du temps, le choix des unités d'information est basé sur l'expérience et l'intuition ce qui implique que plusieurs itérations sont nécessaires pour obtenir des résultats intéressants.

1.1.3. Le filtrage collaboratif et la recommandation

Une troisième méthode suggérée est le filtrage collaboratif. Cette méthode est un moyen efficace d'établir des relations entre différents documents audio (Chen, 2009 ; Congleton et al., 2007). Le modèle le plus simple consiste à permettre à des utilisateurs d'accorder un degré d'appréciation aux pièces musicales. Si plusieurs d'entre eux indiquent apprécier à la fois les pièces musicales x et y alors un lien est créé entre ces pièces. Plus le nombre d'utilisateurs impliqués est grand, plus les résultats obtenus sont intéressants (Linden et al., 2003). Le filtrage collaboratif est à la base de plusieurs algorithmes de recommandation comme celui employé par Amazon³. Cette méthode est toutefois sujette à deux problématiques importantes (Bertin-Mahieux et al., 2010). La première est qu'une pièce nouvellement introduite est généralement peu populaire. Par conséquent, elle est généralement peu recommandée. Ainsi, plusieurs pièces demeurent impopulaires. La seconde problématique est que les pièces populaires ont tendance à devenir plus populaires étant donné qu'elles le sont déjà.

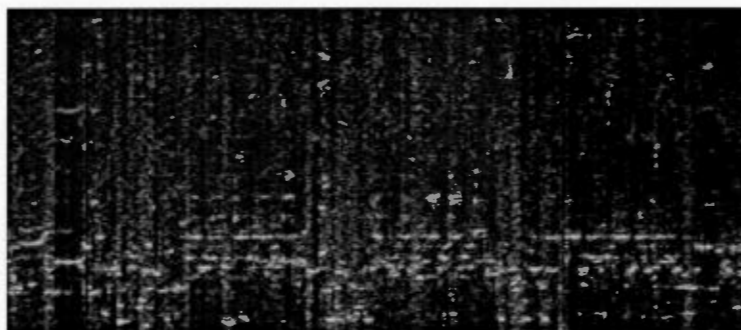
³ <http://www.amazon.com/>

1.1.4. Le fingerprinting

La signature fréquentielle d'une onde sonore est unique pour chacune des pièces musicales. Cette particularité a permis la création de plusieurs systèmes de reconnaissance. Les fréquences dominantes contenues dans le signal sont ciblées et utilisées pour créer une empreinte fréquentielle propre au signal. Cette technique est notamment employée pour identifier des pièces musicales à partir d'enregistrements bruités créés à l'aide d'appareils mobiles.

L'application Shazam est un exemple de l'un de ces systèmes de reconnaissance (Wang, 2003). Shazam est en mesure de reconnaître un grand nombre de pièces musicales. Pour ce faire, l'application a besoin d'analyser un extrait de quelques secondes seulement. L'algorithme utilisé par Shazam exploite le spectrogramme des pièces musicales (figure 1.3). Le spectrogramme est obtenu à l'aide d'une transformée de Fourier⁴. Les pics d'énergie (illustrés par les points pâles dans la figure 1.3) dégagés à l'aide de cette opération sont employés pour former une empreinte fréquentielle (figure 1.4). Les fréquences représentées par des pics d'énergie ainsi que les distances qui les séparent de leurs voisins sont sauvegardées dans une table de hachage pour former des index. Pour être reconnue, une pièce doit être préalablement indexée c'est-à-dire que son empreinte fréquentielle doit avoir été stockée. Lors du processus de reconnaissance, une empreinte est créée pour l'extrait audio utilisé comme critère de recherche. Cette empreinte est ensuite comparée à celles qui sont indexées. Si la pièce recherchée est indexée, le titre et le nom de l'artiste sont retournés.

⁴ La transformée de Fourier est abordée au chapitre 2



1.3. Spectrogramme d'une pièce musicale

L'abscisse correspond au temps et l'ordonnée aux fréquences. Une troisième dimension est exprimée par la couleur. Les points pâles représentent des pics d'énergie. Image capturée d'un prototype développé en C#.



1.4. Empreinte fréquentielle d'une pièce musicale

Empreinte fréquentielle de la pièce illustrée à la figure 1.3. Les pics d'énergie sont utilisés pour créer l'empreinte.

Les méthodes basées sur les empreintes fréquentielles demeurent robustes même lorsque le signal est bruité. Cependant, elles ne sont pas en mesure de reconnaître différentes versions d'une même pièce (Grosche et al., 2012). Selon Grosche, la raison est que les algorithmes utilisés ne sont pas conçus pour tolérer une importante distorsion temporelle non linéaire pouvant affecter le tempo par exemple. Les outils qui exploitent l'empreinte fréquentielle sont en fait des outils de reconnaissance. Par conséquent, ils ne peuvent être utilisés seuls pour effectuer une exploration profonde de données sonores. Ils doivent être jumelés à d'autres mécanismes.

1.2. État de la situation actuelle

À ce jour, il n'existe pas de méthode qui fonctionne dans tous les scénarios. Les méthodes citées précédemment couvrent uniquement certains aspects du traitement de la musique. Elles proposent des stratégies spécifiques que ce soit pour favoriser la recherche, organiser les données, recommander un document ou reconnaître une musique. L'annotation favorise le rappel. Elle est fondée sur l'ajout de descripteurs textuels. Lorsque les métadonnées sont adéquates, elles favorisent le rappel. Toutefois, lorsqu'elles sont incohérentes ou trop subjectives la stratégie est inefficace. De plus, les métadonnées utiles pour une recherche sont souvent inutiles dans le cadre d'autres recherches (Wold et al., 1996). La classification automatisée permet d'organiser les documents audio. Cette opération repose sur une analyse statistique d'unités d'information. Ces unités d'information traduisent certaines caractéristiques dominantes du signal sonore. Cependant, il n'existe aucune garantie que le choix des unités d'information sélectionnées soit optimal. De plus, l'opacité des réseaux de neurones artificiels fréquemment utilisés pour accomplir la classification rend difficile l'interprétation de la contribution des différentes unités

d'information. Le filtrage collaboratif permet de définir des liens entre différents documents. Les préférences des utilisateurs sont exploitées pour établir ces liens. Les liens produits sont pertinents uniquement dans la mesure où les préférences de plusieurs utilisateurs concordent. Finalement, l'empreinte fréquentielle permet de reconnaître une pièce musicale. Par contre, elle ne permet pas une exploration profonde des données.

En plus d'être spécifique à l'exécution de tâches précises, les outils de recherche de similarités musicales actuels sont sujets à diverses limitations. Par conséquent, l'exploration d'une vaste collection de données musicales nécessite un amalgame d'outils complémentaires d'où le besoin de développer une solution permettant d'unifier certaines méthodes de recherche de similarités musicales pour favoriser l'exploration de données musicales.

1.3. Solution proposée

Malgré les progrès réalisés en matière de recherche et de traitement de l'information, l'exploration de données musicales demeure difficile, car il se crée et se diffuse quotidiennement plus de données qu'il nous est possible d'en traiter. Néanmoins, une quantité intéressante de données peut être explorée en utilisant des outils logiciels permettant d'accélérer la consultation et l'analyse des données. Cependant, un problème important demeure, les outils actuels se spécialisent généralement dans le traitement d'un aspect particulier de la musique ce qui limite leur capacité d'exploration. L'objectif général de cette thèse est le suivant : **démontrer la flexibilité et la portabilité des règles d'association dans un contexte d'aide à l'exploration multidimensionnelle de données musicales**. Cet objectif se déploie en trois hypothèses.

1.3.1. Hypothèses

La première hypothèse est que ***l'expérience que procure l'écoute d'une musique dépend fortement du traitement de plusieurs types d'informations dont le niveau d'abstraction varie***. Les traitements qui permettent d'assigner un niveau d'appréciation à une musique ou de mesurer le degré de similarités de deux pièces musicales sont multiples. L'onde sonore est d'abord décomposée et analysée afin de créer des objets sonores. Ensuite, les objets sonores sont manipulés de manière à construire des dimensions qui sont considérées pour créer des classes de similarités (p. ex. les genres musicaux). Finalement, nos préférences musicales, nos expériences personnelles et nos intentions interviennent pour influencer la manière dont la musique est perçue. Chacun des traitements opère sur des unités d'information qui appartiennent à des niveaux d'abstraction différents ce qui laisse présager que les outils d'exploration de données musicales peuvent être améliorés en considérant ces divers niveaux d'abstraction. Il est supposé que chacun des niveaux d'abstraction contribue de manière différente à définir comment la musique est perçue. La portée des travaux en recherche de similarité musicale peut être limitée par le fait que ces travaux portent généralement sur un ou deux de ces niveaux mais rarement sur l'ensemble.

La seconde hypothèse que ***les relations entre les descripteurs utilisés pour décrire une musique sont tout aussi importantes que la présence ou l'absence de ces descripteurs***. Il est soutenu que dans certains cas les relations entre les descripteurs représentent une source d'information encore plus riche. Il est fréquent que de mêmes descripteurs soient utilisés pour décrire un vaste ensemble hétérogène de musiques. La présence seule de ces descripteurs est peu porteuse d'information. Par conséquent, son utilité pour soutenir

l'exploration de données musicales est limitée. Par contre, les relations qu'entretiennent ces descripteurs avec d'autres peuvent mettre en lumière des particularités propres à certaines musiques. Il est donc approprié d'envisager de les utiliser pour orienter l'exploration de données musicales.

La troisième hypothèse est que ***l'utilisation de règles d'association peut d'une part faciliter la mise en œuvre de diverses stratégies utiles à l'exploration de données musicales et d'autre part servir à traiter des informations provenant de différents niveaux d'abstraction.*** L'extraction de règles d'association est une technique employée dans le domaine du forage de données (Amir et al., 2005; Agrawal et al., 1993; Agrawal et Srikant, 1994). L'efficacité des règles d'associations a été démontrée pour le traitement de données textuelles (Biskri et al., 2010; Amir et al., 2005) et de la musique (Tada et al., 2010; Conklin, 2009 ; Arjannikov et Zhang, 2014). Ces méthodes de forage de données ont la capacité de dégager des régularités enfouies à l'intérieur de vaste ensemble de données. De plus, elles peuvent être appliquées à des données dont la nature varie dans la mesure où ces données peuvent être subdivisées en sous-ensembles. Ainsi, la même technique peut être appliquée pour traiter les descripteurs de tous les niveaux d'abstraction et par la même occasion être adaptée à la mise en œuvre de différentes stratégies pouvant soutenir l'exploration de données musicales.

Sur le plan informatique, cette thèse défend l'usage d'une architecture modulaire. Une architecture modulaire est favorable au développement d'un système d'analyse multiniveau. Une architecture modulaire permet de bâtir ou d'adapter rapidement plusieurs chaînes de traitement (Meunier et al, 2005). Comme il n'existe pas de théorie pour guider le choix de l'unité d'information à

privilégier, le développement d'un prototype d'un système d'aide à l'exploration de données musicales peut être facilité par la réutilisation de composantes.

1.4. Contribution

La contribution originale de cette thèse est la démonstration de la portabilité et de la flexibilité des règles d'association dans un contexte d'exploration de données musicales. Les travaux présentés dans cette thèse permettent de faire un pas de plus vers le développement d'une approche générique pour le traitement de données musicales capable de manipuler des descripteurs dont la nature peut varier. Plus précisément, la contribution de cette thèse est :

- Une évaluation exhaustive de la cohésion de trois schémas de classification largement cités dans la littérature ;
- La présentation d'une architecture modulaire et générique centrée sur l'utilisation des règles d'association. Cette architecture a comme particularité de mettre en valeur la flexibilité et la portabilité des règles d'association.
- Le développement et l'évaluation d'une méthode non supervisée de recherche de similarité entre différentes pièces musicales. Contrairement aux méthodes actuelles, la méthode proposée n'est pas opaque. Par conséquent, elle facilite l'identification des unités d'information descriptives ;

- Le développement et l'évaluation d'une méthode d'analyse de la structure des pièces musicales. La méthode proposée, quoique générique, est en mesure de faciliter le repérage des refrains dans des pièces musicales ;
- Le développement et l'évaluation d'une méthode capable de bonifier une liste d'artistes similaires à un artiste cible. Une solution pour réduire le risque d'explosion combinatoire est également exposée.

Des descripteurs appartenant à différents niveaux d'abstraction ont été utilisés lors des expérimentations effectuées dans le cadre de ce travail de recherche. L'aspect original de cette thèse est que toutes les solutions apportées exploitent la même mécanique de base. La même approche a été adaptée afin de couvrir plusieurs problèmes dont est sujet l'exploration des données musicales.

1.5. Structure du document

Le chapitre 2 introduit les notions fondamentales associées aux différentes formes de représentation de la musique. La nature des descripteurs utilisés pour représenter le contenu musical y est présentée. Les descripteurs seront situés par rapport à leur contribution au développement de l'expérience que procure l'écoute d'une musique. En outre, la manière dont les descripteurs participent à la création des différentes perspectives avec lesquelles la musique peut être abordée sera présentée. Certains descripteurs de bas niveau utilisés pour décrire la forme de l'onde sonore seront présentés. La relation entre ces descripteurs et la manière dont le son est perçu sera exposée. Des descripteurs intermédiaires seront également présentés. Les descripteurs intermédiaires

sont utilisés pour représenter des concepts tels que le genre musical. Certains de ces descripteurs résultent de convention. Il sera démontré que l'adoption de ces conventions ne fait pas l'unanimité. En plus des descripteurs de bas niveau et les descripteurs de niveau intermédiaire, des descripteurs de haut niveau seront introduits. Ces descripteurs sont des indices de la valeur accordée par un individu à une musique. Finalement, des mécanismes dédiés à l'extraction de tous ces descripteurs seront exposés. Les mécanismes présentés constituent des mécanismes incontournables pour le traitement automatisé de la musique.

Le chapitre 3 est consacré à la classification de la musique. La classification est l'une des stratégies les plus répandues pour soutenir l'exploration de données musicales. La classification sera positionnée en tant que prétraitement facilitant l'interprétation et la consultation des données musicales. La notion de similarité est une notion clé dans le traitement de données musicales (Martin, 2012). Les étapes du processus qui mène à la création de classes de similarité seront détaillées dans ce chapitre. Quelques méthodes de classification seront également présentées. Les limitations par rapport à ces méthodes seront finalement exposées.

L'usage des règles d'association est au centre des travaux de recherche effectués dans le cadre de cette thèse. Ainsi, le chapitre 4 est entièrement consacré à la présentation des règles d'association. Les fondements théoriques de cette technique de forage de données seront présentés. En particulier, les notions de transaction, d'item et d'itemset seront abordées. La définition générique de la notion de transaction sera précisée. Cette précision est importante dans le cadre des travaux rendus, car c'est ce qui permet l'usage des règles d'association pour extraire des informations cachées au sein des descripteurs des différents niveaux d'abstraction. Les mesures utilisées lors de l'extraction des règles

d'association seront également couvertes. L'algorithme APRIORI sera introduit. Cet algorithme est considéré comme étant la référence dans le domaine de l'extraction des règles d'association. Finalement, des travaux récents portant sur l'usage des règles d'association pour traiter la musique seront sommairement dévoilés afin d'accroître l'intérêt pour cette technique dans un contexte d'exploration de données musicales.

Une architecture pour soutenir l'exploration des données musicales est présentée au chapitre 5. Même si le système développé dans le cadre de cette thèse demeure à ce jour un prototype, les considérations architecturales à retenir pour favoriser son évolution sont données. Les spécifications techniques du prototype développé sont également listées au chapitre 5.

Le chapitre 6 est consacré à la présentation des expérimentations réalisées pour vérifier le potentiel des règles d'association dans un contexte d'exploration de données musicales. Une évaluation manuelle de la cohésion de trois schémas de classification est d'abord présentée. L'objectif de cette évaluation est de démontrer certaines limites liées à l'usage d'un seul type de descripteur. De plus, à partir de cette expérimentation des balises seront données afin de pouvoir comparer les résultats générés à l'aide des règles d'association avec ceux cités dans la littérature. La deuxième expérimentation qui sera présentée dans ce chapitre porte sur l'usage des règles d'association pour définir un degré de similarité entre différentes pièces musicales. Cette expérimentation vérifie si les relations de cooccurrence exprimées par les règles d'association sont suffisamment descriptives pour dégager des relations pertinentes entre des pièces musicales. La troisième expérimentation exposée vise à confirmer la capacité des règles d'association à faciliter l'analyse de la structure des pièces musicales. Entre autres, la manière dont les règles d'association peuvent être

utilisées pour aider à localiser le refrain dans une pièce sera présentée. La troisième expérimentation qui sera présentée est consacrée à la manipulation de descripteurs intermédiaires. Cette expérimentation dévoile comment les règles d'association peuvent être utilisées pour bonifier une liste d'artistes similaires à un artiste cible. Une discussion par rapport aux résultats obtenus lors des différentes expérimentations conclut ce chapitre.

Finalement, la conclusion et les perspectives à venir sont données au chapitre 7.

CHAPITRE II

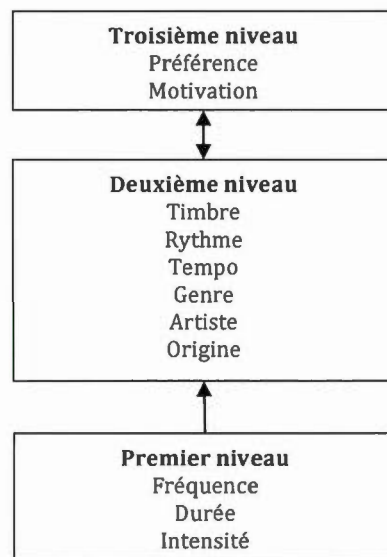
LA MUSIQUE ET SES FORMES DE REPRÉSENTATION

La musique repose sur une organisation séquentielle de sons qui se manifeste dans le temps (Springer et Deutsch, 2001). À partir d'un nombre fini de sons, un nombre infini de mélodies complexes peut être généré (Hoch et al., 2008). Au-delà de cette description, il est difficile de définir précisément ce qu'est la musique. Les frontières qui séparent bruit, mélodie, cacophonie et parole et sont volatiles. Par exemple, certaines musiques expérimentales auront tendance à être considérées davantage comme du bruit que comme une musique. Ce constat amène son lot de difficultés lorsqu'il est question d'exploration de données musicales. Les travaux de Levitin ont démontré que l'écoute de la musique provoque une activité cérébrale que l'écoute de bruits n'engendre pas (Levitin, 2006). Cette observation suggère que la musique est une forme particulière de son.

2.1. Les descripteurs

Tout comme il existe de multiples manières de décrire une musique, il existe plus d'une manière de décrire l'expérience que procure l'écoute d'une musique. L'écoute d'une musique engendre une expérience musicale. L'expérience musicale peut être présentée comme étant l'accomplissement d'un ensemble de jugements par rapport à une musique. Chaque jugement découle d'une analyse

de descripteurs dont le niveau d'abstraction diffère. Ainsi, c'est par l'entremise de l'analyse de ces descripteurs qu'il est possible d'accorder un seuil d'appréciation ou de mesurer la similarité entre deux pièces musicales. Plusieurs évidences suggèrent que l'expérience musicale se crée à travers divers traitements de l'information. Il est possible de grouper ces traitements en fonction de la perspective qu'ils abordent.



2.1. Niveaux d'abstraction des descripteurs utilisés pour décrire la musique

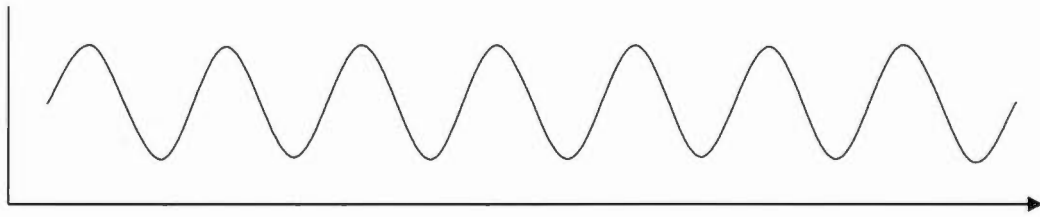
Un niveau d'abstraction est un ensemble de descripteurs informatifs. La contribution d'une unité d'information dans le développement de l'expérience musicale détermine à quel niveau d'abstraction elle appartient. Le premier niveau réfère aux informations de bas niveau. Typiquement, ce sont les propriétés de l'onde sonore. Les traitements appliqués à ce niveau consistent à

grouper certaines propriétés de l'onde sonore afin de former des objets sonores. Le second niveau correspond à la création des dimensions musicales. Les objets sonores sont manipulés pour créer des concepts tels que le genre musical. Des étiquettes sont alors apposées à des séquences particulières de sons. Finalement, le dernier niveau fixe l'appréciation accordée à une expérience musicale.

2.1.1. Les descripteurs de bas niveau

Le son est un phénomène perceptible. Il est produit par des variations de la pression de l'air. Lorsqu'un objet se déplace, il comprime et raréfie successivement les molécules d'air qui l'entoure. Ces variations de la pression de l'air se propagent dans l'environnement. C'est ce qui forme l'onde sonore. Le rôle d'un haut-parleur est de reproduire ces variations. Une séquence de sons devient une musique parce qu'il existe des mécanismes cérébraux chargés de transformer cette séquence en musique. Ainsi, la perception de la musique dépend de mécanismes cognitifs spécifiques chargés de transposer des vibrations en une expérience riche et significative (Patel, 2003). Les descripteurs de bas niveau expriment la forme de l'onde sonore.

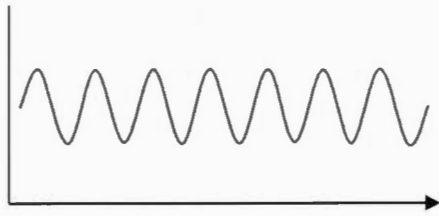
L'onde sonore peut être pure (figure 2.2) ou complexe (figure 2.4). Une onde pure est périodique. Elle possède une forme sinusoïdale qui se répète selon un intervalle de temps fixe nommé période. La vitesse de cette répétition correspond à la fréquence de l'onde sonore.



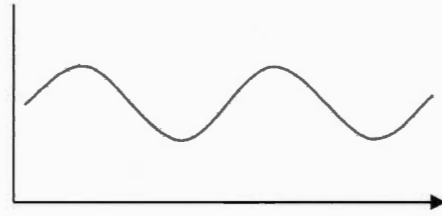
2.2. Illustration d'une onde sonore pure

L'onde sonore pure a une forme sinusoïdale qui se répète selon un intervalle de temps fixe.

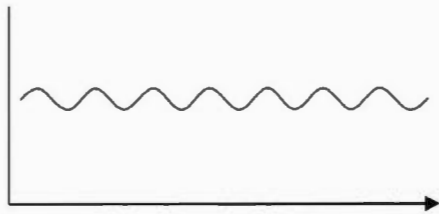
La fréquence détermine la hauteur d'un son. Une fréquence rapide est perçue comme un son aigu (haut) tandis qu'une fréquence lente est perçue comme un son grave (bas). La fréquence est mesurée en hertz (Hz). L'humain perçoit généralement les sons dont la fréquence est comprise entre 20 Hz et 20 000 Hz. Cependant, cet intervalle décroît avec l'âge. Une autre propriété importante de l'onde sonore est son amplitude. L'amplitude détermine l'intensité du son. Plus l'amplitude est importante, plus le son perçu est fort.



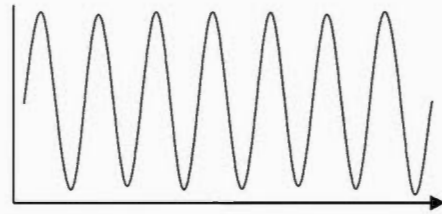
a) Fréquence rapide : son aigu



b) Fréquence lente : son grave



c) Faible amplitude : son faible

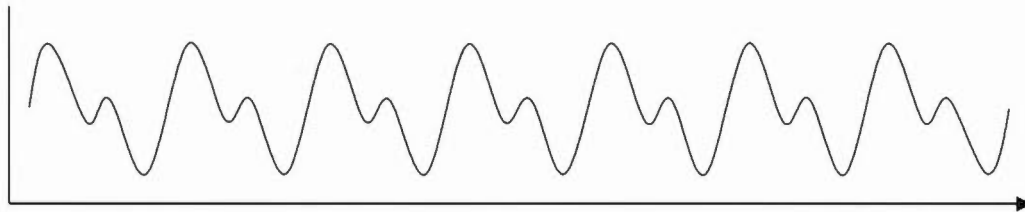


d) Forte amplitude : son fort

2.3. Propriétés de l'onde sonore

La perception que nous avons d'un son dépend entre autres de la fréquence et de l'amplitude de l'onde sonore qui véhicule ce son. Lorsque la fréquence de l'onde qui porte le son est rapide, le son est perçu comme étant aigu. Inversement, lorsque la fréquence de l'onde qui porte le son est lente, le son est perçu comme étant grave. Dans le même ordre d'idée, une fréquence de faible amplitude génère un son faible tandis qu'une fréquence de forte amplitude génère un son fort.

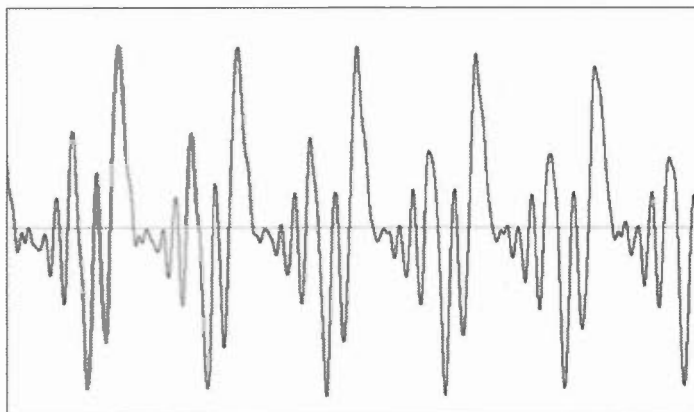
Les ondes sonores possèdent rarement une seule fréquence. Généralement, elles sont composées de plusieurs signaux de fréquence et d'amplitude différente (figure 2.4). Ce sont des combinaisons d'ondes de fréquence et d'amplitude particulières qui accordent aux instruments de musique et à la voix leur sonorité unique (le timbre).



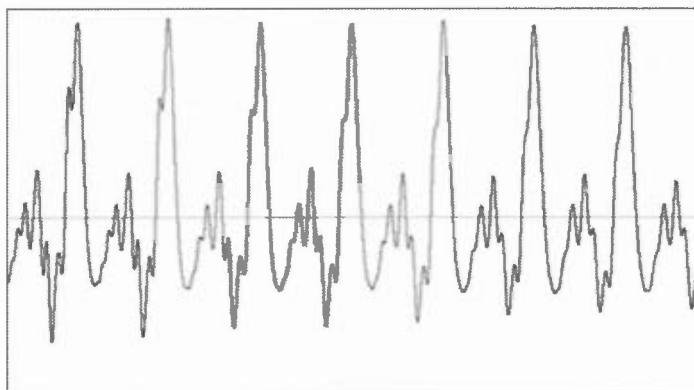
2.4. Onde sonore complexe

L'onde sonore complexe illustrée à la figure 2.4 est composée de deux signaux de fréquence et d'amplitude différente. Chaque instrument introduit des fréquences particulières à l'onde sonore qu'il propage. Ce sont ces fréquences qui accordent aux instruments de musique leur sonorité unique. Cette texture est appelée le timbre.

La figure 2.5 illustre l'onde sonore produite lorsque la note E est jouée à la guitare tandis que la figure 2.6 illustre l'onde sonore produite lorsque cette même note est jouée à la basse. On remarque que même si la même note est jouée, la forme de l'onde diffère d'un instrument à l'autre.



2.5. Onde sonore de la note E (mi) jouée à la guitare



2.6. Onde sonore de la note E (mi) jouée à la basse

Notre système auditif est sans cesse stimulé simultanément par des sons qui proviennent de plusieurs sources différentes. Lorsque des sons sont émis en même temps et à pareille distance ils parviennent à l'oreille au même moment. C'est le cas notamment pour la musique. Les ondes sonores produites par les différents instruments sont mixées pour former une onde sonore complexe. En plus de contenir de l'information sur la mélodie, le rythme, le tempo, etc., l'onde sonore véhicule de l'information sur les instruments utilisés. Bien que l'information soit compressée et canalisée dans un même signal, nous sommes ordinairement en mesure de décompresser et d'assimiler cette information (Levitin, 2006). À ce jour, aucun ordinateur n'est en mesure d'effectuer parfaitement cette tâche.

2.1.2. Les descripteurs intermédiaires

Lors de l'écoute d'une musique, nous sommes stimulés par certaines caractéristiques de l'onde sonore. Chacune de ces caractéristiques contribue à la création d'une ou plusieurs dimensions (Pampalk et al., 2003). La mélodie, le rythme et le timbre représentent quelques une de ces dimensions (Levitin et Tirovolas, 2010). Selon Patel, les mécanismes cognitifs impliqués dans le traitement de ces dimensions sont nombreux (Patel, 2003). Ces mécanismes sont responsables d'extraire les caractéristiques importantes de l'onde sonore et de leur accorder une signification. Plusieurs travaux dédiés à la classification automatisée de la musique dégagent des évidences en faveur de cette hypothèse. Dixon et ses collègues affirment que le rythme peut être utilisé pour catégoriser la musique (Dixon et al., 2003). Chen et Chen ont quant à eux démontré que la catégorisation de la musique peut être réalisée en considérant uniquement le timbre (Chen et Chen, 2009). Ces travaux suggèrent que le rythme et le timbre

nous permettent de catégoriser la musique. Cependant, le rythme et le timbre ne sont pas les seules dimensions pouvant être impliquées dans ce processus (Wold et al., 1996). Le fait d'écouter seulement le rythme d'une musique ou de porter son attention uniquement à la texture des instruments ne permet pas d'obtenir la même expérience que d'écouter une musique dans son ensemble. En fait, plusieurs travaux démontrent que l'efficacité des systèmes de classification automatisée est améliorée lorsque plus d'une dimension est considérée pour catégoriser la musique (Tzanetakis et Cook, 2002, Mitra et Chen, 2007). En somme, les caractéristiques de l'onde sonore engendrent des dimensions et l'analyse de ces dimensions conduit à des indices pouvant être utilisés pour décrire une expérience musicale. Une évaluation des dimensions peut également servir à déterminer un niveau de ressemblance entre deux pièces musicales.

Le genre est utilisé naturellement pour décrire une musique. Plusieurs grandes collections musicales sont structurées en fonction du genre. C'est le cas notamment des collections d'iTunes et d'Archambault⁵ deux détaillants de musique. Dans une certaine limite, l'extraction et l'analyse des caractéristiques de l'onde sonore peuvent mener à l'identification du genre musical. Cependant, les genres musicaux forment des catégories mal définies. L'utilisation du genre pour définir une musique implique la reconnaissance et l'adoption de plusieurs définitions. Cette contrainte amène de nombreuses limitations. Par exemple, un individu n'est pas porté à exploiter un genre avec lequel il n'est pas familier (Rentfrow et al., 2011). Nos préférences musicales influencent la manière dont on conçoit un genre musical. Le genre musical attribué à une musique peut sembler représentatif pour un individu et aberrant pour un autre. Il n'existe pas de bonnes ou mauvaises définitions d'un genre musical. L'étiquetage d'une musique implique une convention sociale. La nomination culturelle est un

⁵ <http://www.archambault.ca>

élément important de la musique. Les étiquettes qui sont apposées et acceptées socialement participent à créer une dimension culturelle à la musique. Cependant, les définitions des genres musicaux ne font pas l'unanimité. Le tableau 2.1 donne les genres attribués à quelques albums par iTunes et Archambault en 2012 tandis que le tableau 2.2 donne ces mêmes informations mais en 2015. Ces tableaux illustrent la subjectivité du genre musical. Les deux détaillants utilisent des définitions différentes. De plus, ces définitions ont évolué en l'espace de 3 ans. Ce manque de régularité suggère qu'utiliser uniquement le genre musical pour décrire une musique n'est pas suffisant.

Artiste	Album	iTunes	Archambault
Arcade Fire	The Suburbs	Alternatif	Pop / Rock
Daniel Bélanger	Nous	Francophone	Francophone
Joe Dassin	Le Jardin du Luxembourg	Francophone	Francophone
Karkwa	Les chemins de verre	Francophone	Francophone
Malajube	La caverne	Francophone	Francophone
Metallica	And Justice for All...	Rock	Métal
Mogwai	Rock Action	Alternatif	Pop / Rock
Sam Roberts Band	Collider	Rock	Pop
The Beatles	White Album	Rock	Pop / Rock
Wilco	Summerteeth	Rock	Pop / Rock

2.1. Genres attribués par iTunes et Archambault.ca en 2012

En examinant les données du tableau 2.1, il est possible de constater que les deux détaillants n'assignent pas le même genre musical aux albums. En effet, iTunes distribue les 10 albums au sein de 3 genres musicaux (alternatif, francophone et rock) tandis qu'Archambault distribue ces mêmes 10 pièces au sein de 4 genres (pop / rock, francophone, métal et pop). Même s'il semble avoir des ressemblances entre les schémas de classification employés par les deux détaillants, seule la définition du genre francophone semble faire l'unanimité en 2012.

Artiste	Album	iTunes	Archambault
Arcade Fire	The Suburbs	Alternatif	Pop / Rock
Daniel Bélanger	Nous	Francophone	Pop (Franco)
Joe Dassin	Le Jardin du Luxembourg	Francophone	Francophone
Karkwa	Les chemins de verre	Francophone	Alternatif
Malajube	La caverne	Alternatif	Francophone
Metallica	And Justice for All...	Rock	Rock
Mogwai	Rock Action	Alternatif	Alternatif
Sam Roberts Band	Collider	Rock	Pop / Rock
The Beatles	White Album	Rock	Pop / Rock
Wilco	Summerteeth	Rock	Rock

2.2. Genres attribués par iTunes et Archambault.ca en 2015

Les schémas de classification utilisés par les deux détaillants ont évolué en 3 ans. Ce changement est particulièrement notable pour le schéma employé par Archambault.

Les données audio sont fréquemment accompagnées de données périphériques. Ces données prennent généralement la forme d'étiquettes textuelles. En plus du genre musical, ces étiquettes renseignent sur le nom de l'artiste, le titre de la pièce, la durée, le distributeur, l'année de publication, etc. Ces données forment une source d'information complémentaire qui précise la nature de la pièce musicale.

Genre Artiste

Musique » Musique francophone » Salomé Leclerc

Sous les arbres

Titre de la pièce

Durée Popularité Prix

Année de publication

Distributeur

Recommandation

Note

Note moyenne : ★★★★★ 28 notes Noter cet article : ★★★★★

Nom	Durée	Popularité	Prix
1 Partir ensemble	3:25	★★★★	\$0.99 ACHETER
2 Sous les arbres	4:13	★★★★	\$0.99 ACHETER
3 Garde-moi collée	3:39	★★★★	\$0.99 ACHETER
4 Ne reviens pas	2:16	★★★★	\$0.99 ACHETER
5 Dans la prairie	4:55	★★★★	\$0.99 ACHETER
6 Tourne encore	3:28	★★★★	\$0.99 ACHETER
7 Longue saison	3:42	★★★★	\$0.99 ACHETER
8 Nos jours	4:21	★★★★	\$0.99 ACHETER
9 Caméléon	2:10	★★★★	\$0.99 ACHETER
10 Noir sur fond blanc	4:33	★★★★	\$0.99 ACHETER
11 Volcan	5:03	★★★★	\$0.99 ACHETER

Total : 12 chansons

D'autres ont aussi acheté

Aux alentours Marie-Pierre Arthur

Le désert des solitudes Catherine Major

Les atomes Martin Léon

Hisser haut David Giguère

Une sorcière comme les autres Jorane

De tune à Alex Neve

2.7. Données textuelles périphériques aux données sonores dans iTunes

2.1.3. Les descripteurs de haut niveau

Les descripteurs de haut niveau ne peuvent être extraits de l'onde sonore ni des données périphériques qui accompagnent les enregistrements sonores. Les descripteurs de haut niveau sont des indices de la valeur accordée par un individu à une musique. L'obtention des descripteurs de haut niveau est importante, car l'analyse seule de l'onde sonore ne permet pas de soutenir efficacement l'exploration de données sonores (Pampalk et al., 2003). Ce constat est d'autant plus marqué lorsque l'objet représente plusieurs versions d'une même pièce (Foucard et al., 2010). La musique peut être reconnue par l'humain peu importe la manière dont elle est produite (Levitin, 2006). Par exemple, une mélodie peut être reconnue même si l'octave à laquelle elle est jouée diffère de la version originale, que son rythme est altéré ou qu'elle est produite à partir d'instruments de musique différents. Les systèmes informatiques actuels ont de la difficulté à accomplir cette tâche. Pour faciliter l'exploration de données musicales, plusieurs auteurs ont suggéré l'utilisation de vues. L'utilisateur est alors capable de visualiser les relations et les distinctions entre différentes musiques. Même si cette solution ne permet pas de résoudre le problème de reconnaissance elle permet néanmoins d'orienter l'utilisateur dans son exploration.

L'ajout de différentes vues sur un ensemble de documents sonores favorise l'exploration de ces documents (Pampalk et al., 2003; McKay et Fujinaga, 2008). Chacune des vues complète l'information apportée par les autres vues. Pampalk et ses collègues ont exposé comment chacune des vues contribue de manière différente à la création de classes de similarité. Dans le même ordre d'idée, McKay et Fujinaga ont exploité des cartes autoadaptatives (SOM) pour générer une représentation visuelle des données sonores (McKay et Fujinaga,

2008). Les cartes autoadaptatives structurent automatiquement les données de manière à ce que celles qui partagent des caractéristiques soient organisées proches les unes des autres et que celles qui ne partagent pas de caractéristique soient distancées. Le modèle proposé par McKay et Fujinaga exploite 3 cartes. La première carte est construite en fonction de la périodicité de l'onde sonore. La périodicité est utilisée comme un indicateur du rythme. La seconde carte est créée en fonction du contenu fréquentiel de l'onde sonore. Le contenu fréquentiel est utilisé pour représenter le timbre. Finalement, une carte est créée manuellement en fonction d'un schéma subjectif. Ainsi, la navigation entre les différentes cartes permet de visualiser comment le rythme et le timbre influencent la catégorisation des données sonores, et ce en comparaison avec une classification manuelle des données. Les résultats obtenus indiquent que les classes produites changent considérablement d'une vue à l'autre. McKay et Fujinaga ont démontré que jumeler des descripteurs dont la nature varie améliore les classifications produites par des systèmes automatisés (McKay et Fujinaga, 2008). Leurs expérimentations révèlent des taux de précision de 96,8% et 78,8% lorsque des caractéristiques de l'onde sonore et des descripteurs de niveau supérieur sont considérées comparativement à des taux de 85,5% et 65,1% lorsque seules des caractéristiques de l'onde sont considérées. Ces observations renforcent la position selon laquelle l'analyse seule de l'onde sonore ne permet pas de soutenir efficacement l'exploration de données sonores.

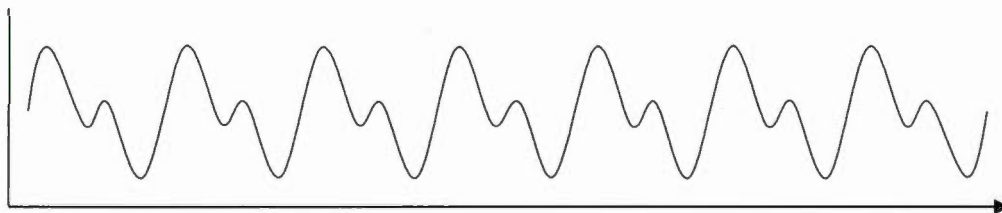
L'expérience musicale est une expérience personnelle. La signification qu'on lui accorde dépend de nos préférences musicales et de nos attentes. Nos préférences proviennent en partie de schémas que nous nous sommes créés depuis l'enfance (Levitin, 2006). La pression sociale modifie néanmoins la manière dont on perçoit la musique. Des stéréotypes sont portés à l'endroit de

certaines genres musicaux. Ces stéréotypes influencent fortement nos préférences musicales (Rentfrow et al., 2011). Finalement, l'effet d'entraînement dirige l'intérêt vers certaines musiques plutôt que d'autres. Par exemple, les adolescents ont tendance à écouter la musique que leurs amis écoutent (Rentfrow et al., 2011). Tous ces facteurs qui contribuent à définir une expérience musicale ne peuvent pas être dégagés uniquement à partir d'une analyse de l'onde sonore. Ces facteurs peuvent être obtenus seulement à partir d'interactions avec l'utilisateur.

Les descripteurs de haut niveau sont des descripteurs qui dressent le profil des utilisateurs (âge, sexe, origine, etc.). Ils peuvent également être constitués des données historiques (p. ex. quelles sont les dernières pièces écoutées) permettant de cibler l'intérêt d'un utilisateur. L'utilisation de descripteurs de haut niveau a comme objectif de personnaliser et raffiner les outils d'exploration de données musicales en fonction de la motivation et des préférences des utilisateurs.

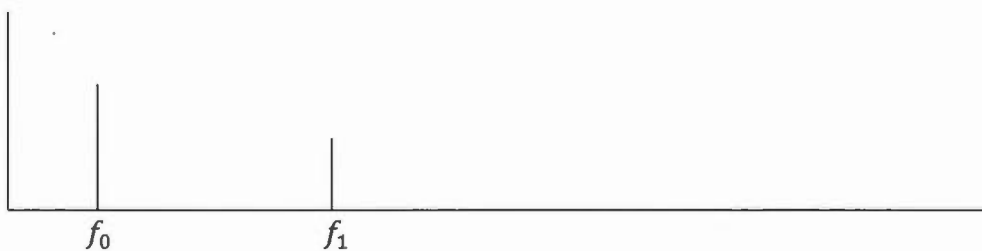
2.2. Le traitement du signal

Plusieurs équations formalisent les propriétés et le comportement de l'onde sonore (Cavicchi, 1999). L'application d'algorithmes rend l'extraction de certaines unités d'information de l'onde sonore possible et relativement facile à réaliser. Il est possible de subdiviser ces algorithmes en deux catégories. La première catégorie représente les algorithmes applicables au domaine spatial (figure 2.8) tandis que la seconde représente les algorithmes applicables au domaine fréquentiel (figure 2.9).



2.8. Forme de l'information du domaine spatiale

Le domaine spatial réfère à la forme de l'onde dans le temps.

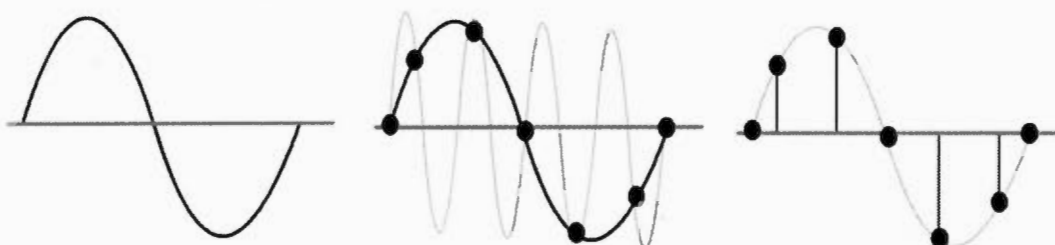


2.9. Forme de l'information du domaine spectral

Le domaine spectral réfère au contenu fréquentiel de l'onde sonore.

2.2.1. Le domaine spatial

Le domaine spatial réfère à la forme de l'onde. La numérisation du son consiste à prélever un nombre fini de points sur l'onde sonore de manière à pouvoir la reconstituer. L'échantillonnage est effectué à un intervalle de temps régulier nommé fréquence d'échantillonnage. Cette fréquence doit être suffisamment élevée afin de préserver la forme de l'onde sonore et ainsi reproduire le son le plus fidèlement possible. Ce procédé est illustré à la figure 2.10.



2.10. Échantillonnage de l'onde sonore

Selon le théorème de Shannon (Shannon, 1948), la fréquence d'échantillonnage doit être au moins égale au double de la plus haute fréquence contenue dans le signal analogique. En deçà de cette limite théorique, il n'est pas possible de reconstituer un signal à partir d'échantillons sans perte d'information. Comme l'oreille humaine perçoit en moyenne les sons situés entre 20 hertz et 20 000 hertz, une fréquence d'échantillonnage de 40 000 hertz est nécessaire pour obtenir un enregistrement de qualité. La norme actuelle pour les documents

audio est de 44 100 hertz. Le domaine spatial est donc représenté par de vastes ensembles de points d'échantillonnage.

2.2.1.1. Les caractéristiques du domaine spatial

L'intérêt pour les caractéristiques liées au domaine spatial réside dans le fait qu'elles sont généralement simples à dégager. En effet, ces mesures peuvent être extraites directement à partir des points d'échantillonnage. Il existe plusieurs caractéristiques dédiées au domaine spatial. Le taux de passage par zéro, la durée logarithmique de l'attaque et le centroïde temporel figurent parmi les caractéristiques du domaine spatial citées dans la littérature consacrée au traitement de la musique (Tzanetakis et al., 2002 ; Lu et al., 2001 ; Herrera, 2001 ; Mckay et Fujinaga, 2008 ; Lewis et Wierzchowska, 2007).

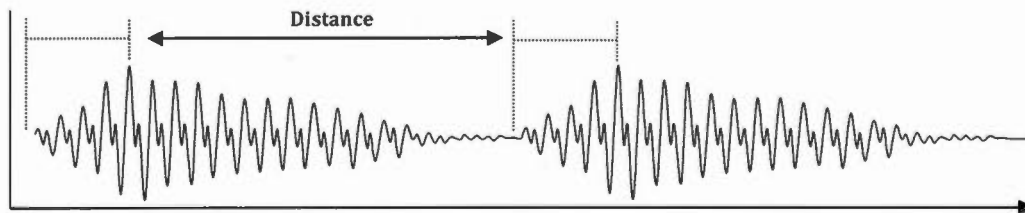
Le taux de passage par zéro (*Zero Crossing Rate* ou *ZCR*) consiste à comptabiliser le nombre de fois que le signal passe par zéro. Le calcul du ZCR est donné par l'équation 2.1 où x représente un vecteur contenant les données brutes et T la durée de l'enregistrement. Le taux de passage par zéro est utilisé notamment pour mesurer le bruit dans le signal (Tzanetakis et al., 2002). Il peut également être utilisé pour estimer le timbre et ainsi aider à différencier la musique d'autres sons (Lu et al., 2001).

$$zcr = \frac{1}{T} \sum_{t=1}^T X_t \text{ où } X_t = \begin{cases} 0, & x_t \cdot x_{t-1} > 0 \\ 1, & x_t \cdot x_{t-1} \leq 0 \end{cases} \quad (2.1)$$

L'attaque désigne l'évolution de l'intensité d'un son du moment où il est émis jusqu'au moment où il atteint son intensité maximum. La figure 2.11 illustre cette propriété. Tout comme le timbre, l'attaque donne des indications sur la nature de la source d'un son. Par exemple, l'attaque peut servir à classifier les sons de percussions (Herrera, 2001). L'équation 2.2 donne la formule du LAT (*Log Attack Time*) fréquemment employé pour désigner l'attaque.

$$lat = \log_{10}(fin\ attaque - debut\ attaque) \quad (2.2)$$

Le temps moyen entre deux attaques peut servir à estimer le rythme (Mckay et Fujinaga, 2008).



2.11. Distance entre deux attaques

Le centroïde temporel est une autre mesure qui s'applique au domaine spatial. Cette mesure permet de situer à quel endroit l'énergie est la plus importante dans le signal. La formule du centroïde temporel est donnée par l'équation 2.3 où X représente un vecteur contenant les données brutes et T la durée de l'enregistrement. Le centroïde temporel peut être utilisé pour différencier des

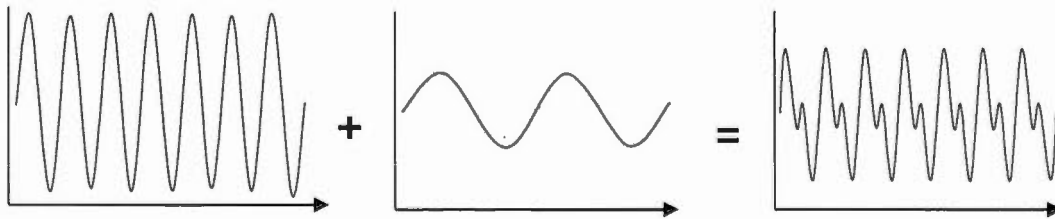
instruments de musique. Par exemple, il peut servir à différencier les notes d'un piano par rapport à celles d'un orgue (Lewis et Wierzchowska, 2007).

$$ct = \frac{\sum_{t=1}^T tX_t}{\sum_{t=1}^T X_t} \quad (2.3)$$

Les caractéristiques liées au domaine spatial sont rarement utilisées seules. Elles sont généralement utilisées conjointement avec des approches liées au domaine spectral.

2.2.2. Le domaine spectral

Le domaine spectral réfère aux fréquences véhiculées par l'onde sonore. Le mathématicien français Joseph Fourier (1768 - 1830) a démontré qu'il était possible d'approximer un signal périodique complexe en fonction de signaux élémentaires de forme sinusoïdale appelés harmoniques.



2.12. Décomposition d'un signal en terme de sinusoides

La décomposition d'un signal périodique en une somme de sinusoïdes est appelée série de Fourier. La forme générale de la série de Fourier pour un signal de période T est donnée à l'équation 2.4 où ω , a_0 , a_n et b_n représentent respectivement la fréquence fondamentale du signal et des scalaires appelés coefficients de Fourier.

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega t) + b_n \sin(n\omega t) \text{ où } \omega = \frac{2\pi}{T} \quad (2.4)$$

Les coefficients de Fourier sont donnés par les équations 2.5, 2.6 et 2.7.

$$a_0 = \frac{1}{T} \int_0^T f(t) dt \quad (2.5)$$

$$a_n = \frac{2}{T} \int_0^T f(t) \cos(n\omega t) dt \quad (2.6)$$

$$b_n = \frac{2}{T} \int_0^T f(t) \sin(n\omega t) dt \quad (2.7)$$

La fréquence fondamentale est la fréquence la plus basse contenue dans un signal. Cette fréquence particulière caractérise la hauteur d'un son. Par conséquent, il est possible de préciser la nature d'un son en analysant sa fréquence fondamentale. À titre d'exemple, il est possible de différencier la voix d'un homme de celle d'une femme ou d'un enfant uniquement en considérant la fréquence fondamentale. En effet, en moyenne la fréquence fondamentale de la voix d'un homme se situe dans un intervalle de 100 et 150 hertz comparativement à un intervalle de 200 à 300 hertz pour une femme et un intervalle 350 à 400 hertz pour un enfant (Caillaud et Leriche, 1999). L'analyse de la fréquence fondamentale peut également mener à l'identification de notes de musique et ultimement à la transcription d'une mélodie (Poliner et al., 2007). Les harmoniques, quant à elles, apportent une texture au son. L'analyse des harmoniques peut mener à la reconnaissance de l'instrument utilisé pour produire une mélodie (figure 2.5 et figure 2.6).

Sur le plan informatique, il est possible de passer du domaine temporel au domaine spectral à l'aide l'algorithme de transformé de Fourier rapide (FFT).

2.2.2.1. Les caractéristiques du domaine spectral

L'analyse spectrale permet l'extraction de plusieurs caractéristiques associées aux fréquences. Le centroïde spectral, la répartition spectrale, les coefficients ceptraux représentent quelques une de ces caractéristiques.

Le centroïde spectral est une caractéristique fréquemment employée dans le domaine spectral. Cette caractéristique est l'équivalent du centroïde temporel mais mesuré dans le domaine spectral. Les travaux de Tzanetakis suggèrent que

cette caractéristique peut contribuer à identifier le genre musical (Tzanetakis, 2002). Le centroïde spectral donne la forme du spectre où les centroïdes ayant la valeur la plus importante représentent les régions contenant le plus de hautes fréquences. La formule du centroïde spectral est donnée par l'équation 2.8 où x est une fréquence observée et $p(x)$ la probabilité que cette fréquence soit observée (2.9) (Peeters, 2004).

$$\mu = \int x p(x) dx \quad (2.8)$$

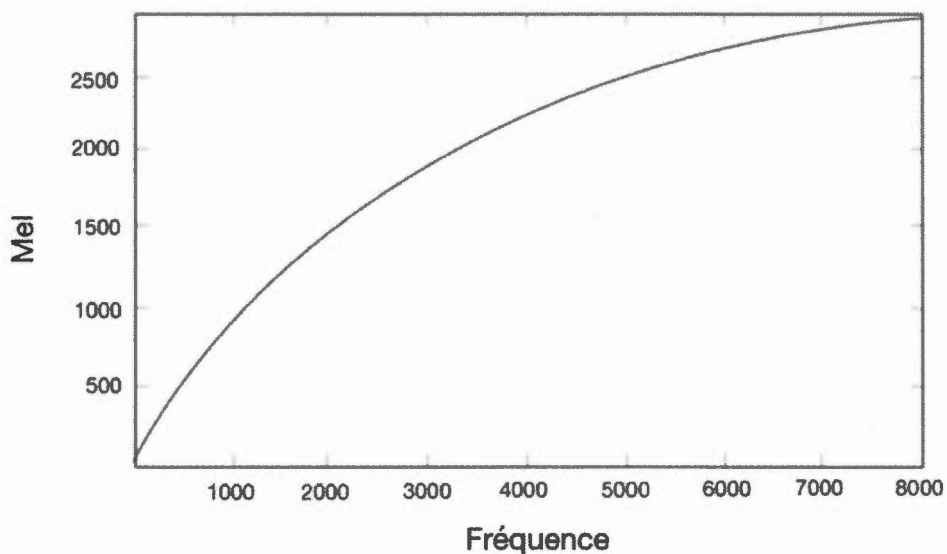
$$p(x) = \frac{\text{amplitude } v(x)}{\sum_{i=0}^{\text{nombre fréquence}} v(x_i)} \quad (2.9)$$

La répartition spectrale est une mesure utilisée conjointement avec le centroïde spectral. Cette mesure donne l'étalement du spectre par rapport au centroïde. Le calcul de la répartition spectrale est donné par l'équation 2.10 (Peeters, 2004).

$$\sigma^2 = \int (x - \mu)^2 p(x) dx \quad (2.10)$$

Les notes de musique réfèrent à des fréquences particulières. Par exemple, fréquence 440 Hz représente la note LA. La différence de hauteur entre deux

notes de musique est nommée intervalle. Un intervalle peut être ascendant (deuxième son plus aigu que le premier) ou descendant (deuxième son plus grave que le premier). On nomme octave l'intervalle qui sépare deux sons dont les fréquences fondamentales ont un rapport de fréquence égal à 2. Ces sons sont perçus par l'oreille de façon comparable et par conséquent ils sont identifiés par la même note. Ainsi, les fréquences 880 Hz, 440 Hz, 220 Hz, 110 Hz et 55 Hz représentent toute la note LA. Cependant, plus la fréquence est élevée, plus cette perception est altérée. En effet, au-delà d'un certain seuil, le changement d'octave n'est plus perçu comme étant lié au doublement de la fréquence. L'échelle de Mel est une échelle de fréquence adaptée à cette distorsion (Stevens et al., 1937). Elle est conçue de manière à ce que 1 000 Mels correspondent à 1 000 Hz. La figure 2.13 illustre la relation entre les Mels et les fréquences.

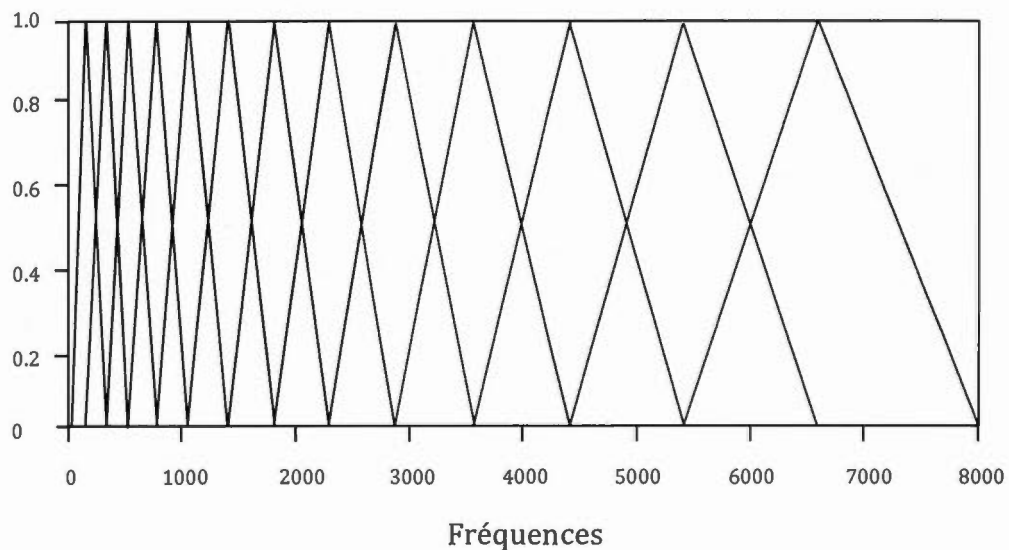


2.13. Courbe de correspondance Mel – Fréquence

La formule permettant de convertir un Mel en Hertz est donnée par l'équation 2.11.

$$mel(f) = 1127.01048 * \ln\left(1 + \frac{f}{700}\right) \quad (2.11)$$

Dans le traitement du signal, l'échelle de Mel est utilisée pour extraire des coefficients cepstraux (MFCC). Sommairement, les MFCC sont obtenus en appliquant un filtre triangulaire à l'échelle de Mel au résultat de la FFT. La figure 2.14 illustre la forme du filtre utilisé pour obtenir les MFCC.



2.14. Filtre triangulaire à l'échelle de Mel

L'avantage des MFCC est qu'il s'agit d'une représentation compacte. L'utilisation d'une dizaine de coefficients est suffisante pour obtenir de bons résultats.

2.2.3. Les librairies dédiées à l'extraction des caractéristiques

L'extraction des propriétés de l'onde sonore intègre diverses notions étroitement liées au traitement du signal. La recherche de nouvelles caractéristiques pour exprimer ces propriétés est un sujet de recherche en soi. Il existe plusieurs caractéristiques et une multitude de stratégies qui exploitent ces caractéristiques. Les caractéristiques présentées précédemment représentent seulement un échantillon de celles existantes. Afin de soutenir la recherche dédiée au traitement automatisé de la musique, plusieurs librairies ont été développées et partagées.

Marsyas⁶ (Music Analysis, Retrieval and Synthesis for Audio Signals) est l'une des premières librairies proposées pour faciliter l'extraction d'informations significatives contenues dans l'onde sonore (Tzanetakis et Cook, 2000 ; Bray et Tzanetakis, 2005). La version 0.4 de la librairie permet l'extraction de 124 propriétés. Elle est l'œuvre d'un groupe de chercheurs dirigés par Tzanetakis. Marsyas est une librairie open-source développée en C++. Cette particularité lui octroie une certaine popularité. Au cours des années, cette librairie a été utilisée autant dans le domaine de la recherche (Schindler et Rauber, 2014 ; Yang et al., 2008) que dans celui de l'industrie (Last.FM, Yahoo, etc.). Depuis le développement de Marsyas, plusieurs autres librairies ont été proposées. Par

⁶ <http://marsyas.info/index.html>

exemple, MIRtoolbox⁷ est une librairie développée pour être utilisée avec Matlab (Lartillot et Toivianen, 2007). MIRtoolbox permet de manipuler 376 propriétés. MK2⁸ est une autre librairie dédiée à l'extraction de propriétés des enregistrements sonores. Cette librairie se distingue par le fait qu'elle offre la possibilité de distribuer les traitements à travers plusieurs ordinateurs (Downie et al., 2004). Le tableau 2.3, adapté de Beveridge et Knox, dresse une liste de quelques-unes de ces boîtes à outils permettant l'extraction de propriétés de l'onde sonore (Beveridge et Knox, 2012).

Boîte à outils	Nombre de propriétés
MIRtoolbox	376
PsySound3	24
Marsyas 0.4	124
Marsyas 0.1	32
Sound Description Toolbox	187
Lu Implementation	71

2.3. Boîtes à outils d'extraction et propriétés couvertes

Tableau adapté de (Beveridge et Knox, 2012)

Les boîtes à outils offrent des mécanismes permettant l'extraction des caractéristiques qui expriment les propriétés de l'onde sonore. L'une des

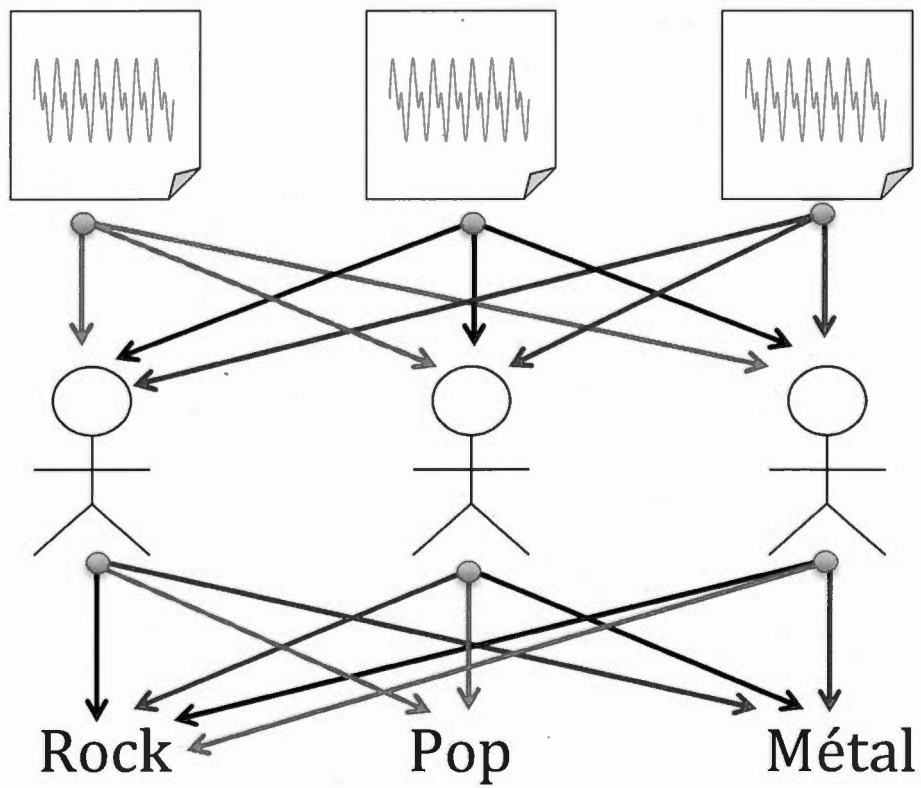
⁷ <https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>

⁸ <http://www.music-ir.org/evaluation/m2k/>

critiques envers ces boîtes à outils est que plusieurs d'entre elles sont peu intuitives. jMIR⁹ est une suite de logiciels développée par McKay et destinée à soutenir la recherche consacrée au MIR (McKay, 2010 ; McEnnis et al., 2005). jMIR est présenté comme un ensemble de logiciels conviviaux. Cet ensemble comprend notamment le module jAudio permettant l'extraction de 28 caractéristiques différentes de l'onde sonore, dont celles présentées précédemment. jMIR facilite également l'extraction d'informations culturelles disponibles sur Internet. Afin d'assurer un certain niveau de portabilité, jMIR est implémenté en Java. L'objectif de cette suite de logiciels est de fournir aux chercheurs un ensemble d'outils leur permettant de concevoir, tester et partager de nouveaux modèles de classification automatique de la musique.

L'usage d'une mesure peut amener à estimer certaines dimensions de la musique. Par exemple, le contenu fréquentiel peut être utilisé pour désigner le timbre. En combinant plusieurs mesures, il est possible de produire des classifications pertinentes. La création de classes de similarité favorise l'exploration musicale, car elle regroupe des pièces qui partagent certains attributs. Les classifications générées automatiquement reposent sur une évaluation de descripteurs objectifs. Même si les classifications produites sont statistiquement cohérentes elles peuvent sembler non significatives pour un utilisateur. L'une des explications est que ces classifications ne tiennent pas compte de critères subjectifs qui varient d'un individu à l'autre.

⁹ <http://jmir.sourceforge.net/>



2.15. Les schémas de classification diffèrent d'un individu à l'autre

2.3. L'acquisition de données subjectives

Une mesure subjective prend en considération à la fois les données et l'utilisateur de ces données (Geng et Hamilton, 2006). Pour définir une telle mesure, un accès aux préférences de l'utilisateur est nécessaire. Cet accès est possible à partir d'interactions avec l'utilisateur ou en lui présentant une représentation de ce que le système sait sur ses attentes. Les systèmes de recommandation constituent un exemple de système qui se sert d'interaction pour raffiner les données présentées (Linden et al., 2003). Les modèles les plus simples se servent uniquement du taux d'appréciation fourni explicitement par l'utilisateur pour adapter les données présentées. Cependant, les modèles de recommandation les plus complexes utilisent des informations comme l'historique des requêtes effectuées et la localisation géographique pour influencer les résultats. L'utilisation de ces données doit toutefois se faire avec précaution, car elles sont parfois volatiles (Linden et al., 2003). L'ajout de nouvelles informations force une mise à jour constante du profil de l'utilisateur. Une mise à jour continuelle peut engendrer un coût computationnel élevé. Néanmoins, l'ajout de mesures subjectives offre une perspective impossible à obtenir à partir d'une analyse objective des propriétés de l'onde sonore.

L'utilisation seule de mesures subjectives à des fins d'exploration musicale n'est pas suffisante, car ces mesures ont tendance à privilégier les relations les plus fréquentes (Bertin-Mahieux et al., 2010). Par conséquent, l'usage de mesures objectives et subjectives pour soutenir l'exploration musicale est à préconiser afin de découvrir des relations cachées et potentiellement intéressantes.

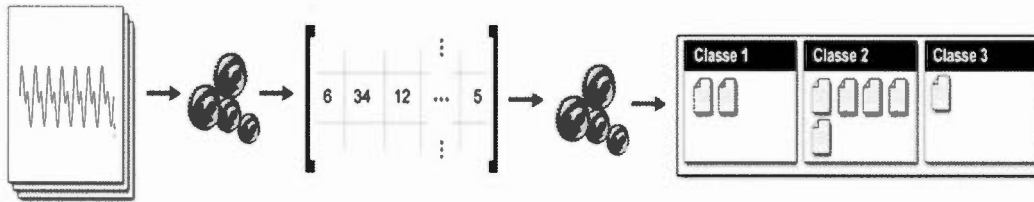
CHAPITRE III

LA CLASSIFICATION

L'une des stratégies les plus communes pour soutenir l'exploration de données sonores consiste à organiser les données de manière à en faciliter l'accès. La classification et le clustering sont deux solutions communément appliquées au problème d'organisation de l'information. La classification a comme objectif d'assigner des données à des classes de similarités prédéfinies tandis que le clustering a comme objectif de partitionner des données de manière à ce que celles qui partagent des similarités soient regroupées. Ces méthodes sont souvent présentées comme des prétraitements facilitant la consultation et par conséquent l'exploration de données. Le processus qui mène à la création automatique de classes de similarités comporte trois étapes importantes :

- 1) La préparation des données ;
- 2) L'application d'une méthode de classification ;
- 3) L'interprétation des résultats.

Les deux premières étapes du processus peuvent être automatisées à l'aide d'algorithmes et de fonctions mathématiques. La troisième étape est plus subjective et nécessite généralement un expert du domaine étudié afin de valider la qualité des classes de similarités obtenues. Ce processus est illustré à la figure 3.1.



3.1. Processus de classification numérique

Illustration tirée de (Rompré et al., 2006).

La pertinence d'une classification est habituellement jugée en fonction de l'homogénéité des classes qui en résulte (Bouveyron, 2006). Ce critère d'évaluation est toutefois relatif. L'évaluation d'une classification est une opération subjective. L'examen d'une classe par un intervenant est accompli à partir de ses objectifs de recherche et de ses connaissances du domaine abordé. Ainsi, une classification cohérente pour certains peut être sans intérêt pour d'autres et vice versa. Cette subjectivité se manifeste dès la création d'une classification. Les connaissances générales et les préférences musicales d'un individu peuvent influencer considérablement sa manière de concevoir un modèle de classification basé sur le genre musical par exemple. Ainsi, deux individus peuvent classer de manière différente un même ensemble de musiques même s'ils considèrent tous les deux le genre comme critère de classification. La qualité recherchée pour un modèle de classification n'est pas d'être en mesure de fournir un schéma universel mais plutôt d'être capable de cibler les informations pertinentes dans un contexte donné et de déterminer comment ces informations peuvent être utilisées pour produire des classes de similarité cohérentes (Estes, 1994).

3.1. La préparation des données

La classification repose sur l'identification et l'évaluation des caractéristiques qui permettent de différencier une classe d'une autre. La présence ou l'absence d'une caractéristique est un indice permettant de cibler la classe à laquelle appartient une entité. La recherche des caractéristiques optimales à des fins de classification est un problème ouvert. Il n'existe pas de solution universelle à ce problème (Biskri et Meunier ; 2002). Prenons par exemple 5 pièces de musique identifiées M1, M2, M3, M4 et M5 et 5 caractéristiques identifiées C1, C2, C3, C4 et C5. Il est possible de décrire ces 5 pièces de musique en fonction de la présence ou de l'absence de ces 5 caractéristiques. Le tableau 3.1 donne les descriptions ainsi créées. Sachant que les 5 pièces appartiennent à une même classe de similarité X, la distribution des caractéristiques au sein de ces pièces est un indice permettant de déterminer ce qui distingue cette classe des autres.

En examinant la distribution des caractéristiques donnée au tableau 3.1, on remarque que la caractéristique C5 est présente dans tous les membres de la classe X. Ainsi, il est possible de déduire que cette caractéristique est un indicateur permettant de reconnaître un membre de la classe X. Supposons maintenant qu'un échantillon plus large est considéré et que dans cet échantillon plusieurs classes soient représentées. Si cette même caractéristique C5 est présente dans toutes les pièces alors son pouvoir discriminant s'estompe et elle ne peut plus être utilisée pour distinguer les membres de la classe X. La plupart du temps, c'est à travers l'expérience que les caractéristiques sont identifiées (Anderson, 1995 ; Estes 1994).

Musique	Caractéristique				
	C1	C2	C3	C4	C5
M1	+	+	+	+	+
M2	+	-	+	-	+
M3	+	+	+	+	+
M4	-	-	+	+	+
M5	+	+	-	+	+

3.1. Répartition de caractéristiques à l'intérieur des membres d'une classe

Le signe + indique la présence de la caractéristique tandis que le signe - indique son absence

Considérons que les caractéristiques retenues soient suffisamment discriminantes pour mener une classification. Un degré de similarité entre deux entités peut être obtenu en vérifiant uniquement la concordance de leurs caractéristiques. Estes (Estes, 1994) soutient que ce degré peut être calculé à partir d'une procédure appelée *règles des produits*. Cette procédure consiste dans un premier temps à assigner la valeur 1 lorsque la caractéristique est présente ou absente pour les deux entités et une valeur d'asymétrie s lorsque ce n'est pas le cas. Ces valeurs sont ensuite multipliées pour obtenir le degré de similarité. Par exemple, selon la méthode d'Estes, le degré de similarité entre les pièces M1 et M2 est égal à s^2 puisque $1 \times s \times 1 \times s \times 1 = s^2$ et le degré de similarité entre les pièces M1 et M3 est égale à 1 puisque $1 \times 1 \times 1 \times 1 \times$

1 = 1. La règle des produits d'Estes peut être généralisée par l'équation 1.1 où N correspond au nombre de caractéristiques considérées et k au nombre de caractéristiques concordantes.

$$\text{sim}(M1, M2) = s^{N-k} \quad (3.1)$$

Les travaux de Rosh et Mervis (Rosh et Mervis, 1975) ont démontré que le degré de similarité d'une entité par rapport à la classe à laquelle elle appartient est fortement corrélé avec le degré de similarité qu'elle entretient avec les autres membres de cette classe. En comparant les entités d'une classe une à une à l'aide de la règle des produits d'Estes puis en additionnant les valeurs obtenues et il est possible d'obtenir le degré d'appartenance d'une entité envers sa classe. Le résultat de cette opération peut être représenté sous la forme d'un tableau comme indiqué au tableau 3.2. Admettons que le facteur d'asymétrie soit fixé à 0.1 alors les degrés d'appartenance de M1, M2, M3, M4 et M5 envers la classe X sont données par le tableau 3.3. Les entités avec le degré d'appartenance le plus élevé sont alors considérées comme étant les prototypes les plus représentatifs de la classe.

	M1	M2	M3	M4	M5	Degré de similarité
M1	1	s^2	1	s^2	s	$2 + s + 2s^2$
M2	s^2	1	s^2	s^2	s^3	$1 + 3s^2 + s^3$
M3	1	s^2	1	s^2	s	$2 + s + 2s^2$
M4	s^2	s^2	s^2	1	s^3	$1 + 3s^2 + s^3$
M5	s	s^3	s	s^3	1	$1 + 2s + 2s^3$

3.2 : Calcul du degré de similarité des membres d'une catégorie.

Le degré de similarité des membres de la catégorie est calculé à l'aide de la règle des produits d'Estes (Estes, 1994).

Exemplaire	Degré d'appartenance $S = 0.1$	Degré d'appartenance relative
M1	2.12	$2.12/7.50 = 0.28$
M2	1.03	$1.03/7.50 = 0.14$
M3	2.12	$2.12/7.50 = 0.28$
M4	1.03	$1.03/7.50 = 0.14$
M5	1.20	$1.20/7.50 = 0.16$

Somme : 7.50

3.3. Degré d'appartenance des membres à leur catégorie

Tableau adapté de (Estes, 1994)

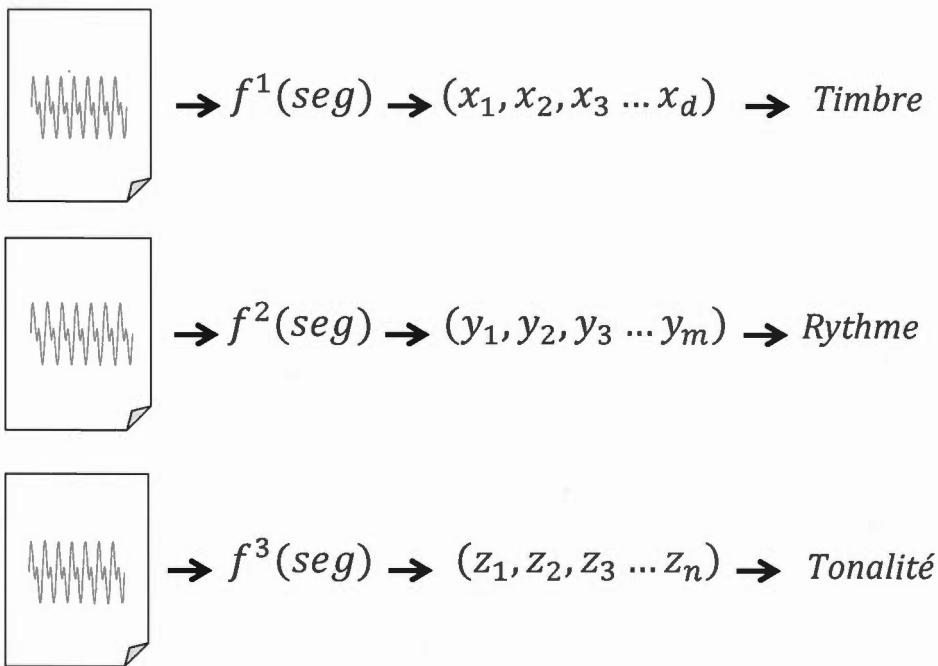
La simplicité du modèle d'Estes le rend intéressant. Ce modèle illustre bien comment la distribution des caractéristiques au sein d'un ensemble de données peut être utilisée à des fins de classification. La plupart des méthodes de classification actuelle reposent sur ce principe général. L'utilisation pratique du modèle d'Estes est cependant limitée par sa grande sensibilité (Eidenberger, 2012).

La plupart des méthodes de classification numérique actuelle opèrent sur des représentations vectorielles des données. Un ensemble de données voué à être classifié peut être exprimé sous la forme d'une matrice $n \times d$ où n correspond au nombre d'entités représentées et d au nombre de caractéristiques utilisées pour décrire ces entités. Ainsi, chaque rangée de cette matrice forme une représentation simplifiée d'une entité (Zaki et Meira, 2014). Dans un contexte d'exploration de données numériques, une entité représente un document, une portion de document, un enregistrement dans une base de données, etc.

$$D = \begin{array}{c|ccc} & \mathbf{u}_1 & \cdots & \mathbf{u}_d \\ \hline \mathbf{e}_1 & x_{11} & \cdots & x_{1d} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{e}_n & x_{1n} & \cdots & x_{nd} \end{array} \quad (3.2)$$

La matrice ci-dessus correspond à une représentation simplifiée d'un ensemble de données D . Chaque rangée e_i correspond à une entité contenue dans D tandis que chaque colonne u_j correspond à une unité d'information utilisée pour représenter une caractéristique de cette entité. Ainsi, les rangés $x_i = (x_1, x_2, x_3 \dots x_d)$ constituent des représentations vectorielles des entités. Les données audio, comme pour la majorité des autres formats de données, ne sont

pas nativement représentées sous cette forme. Il est cependant possible de les transposer à l'aide d'une fonction f . Une telle fonction prend en entrée un sous-ensemble de données et retourne en sortie un vecteur $x_i = (x_1, x_2, x_3 \dots x_d)$. Pour ce faire, les données brutes sont scrutées de manière à pouvoir extraire les attributs utiles pour exprimer les caractéristiques ciblées. La fonction f est souvent spécifique au domaine abordé. La raison est que les caractéristiques à cibler ont tendance à varier d'un domaine à l'autre. Cette particularité s'applique également aux dimensions de la musique. Ainsi, une fonction f est généralement nécessaire pour chacune des dimensions considérées.



3.2. Création des représentations vectorielles

Des fonctions spécifiques à une des dimensions de la musique sont utilisées pour créer des représentations vectorielles des documents sonores. L'input des fonctions est un segment des données brutes.

3.1.1 Les données descriptives

Les caractéristiques qui définissent une entité se présentent rarement de manière explicite dans les données. Il est généralement nécessaire de les déduire à partir d'unités d'information extraites des données brutes. Une telle unité d'information, également appelée unif, doit répondre à 4 contraintes pour assurer un niveau d'adaptabilité intéressant (Biskri et Meunier, 2002) :

- 1) Elle doit être contenue dans les données brutes ;
- 2) Il doit être facile¹⁰ sur le plan informatique de la repérer ;
- 3) Elle doit être statistiquement comparable. Autrement dit, il doit être possible de calculer sa fréquence d'apparition et par conséquent estimer sa distribution et la régularité à laquelle plusieurs unités cooccurrent ;
- 4) Sa nature dépend de la caractéristique à exprimer. Que l'unité d'information soit comprise hors de son contexte n'est pas une contrainte en soi.

Le rôle et l'importance d'une unité d'information dépendent du domaine d'application et de manière plus concise du problème posé et de l'objectif à atteindre (Kunt, 2000). Conséquemment, le choix d'une unité d'information peut restreindre le domaine d'application. Or, l'extraction des unités d'information se heurte à une difficulté théorique d'envergure : il n'existe pas de théorie pour guider le choix des unités d'information à considérer lors de la résolution d'un problème (Kunt, 2000). Cette difficulté théorique est accentuée dû au fait que la

¹⁰ Du moins, il doit être possible de les repérer.

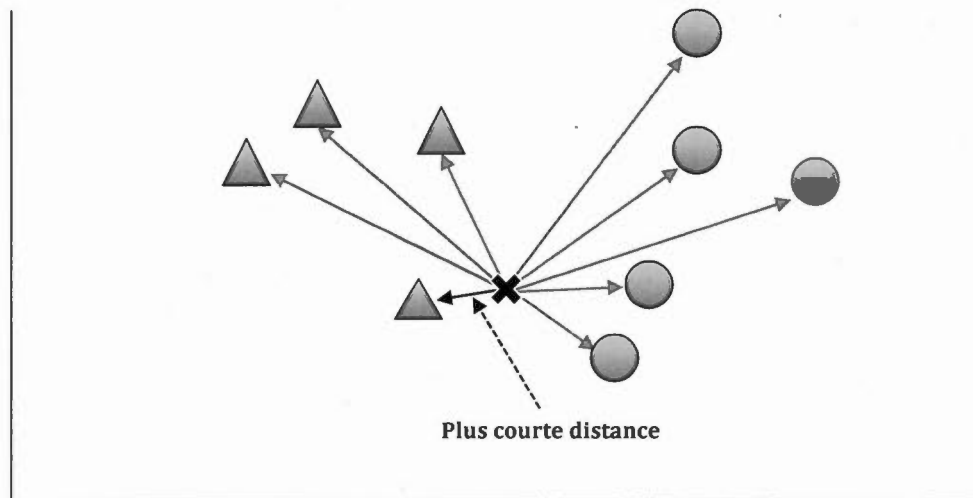
musique possède plusieurs dimensions. Le choix de l'unité d'information repose principalement sur des intuitions et est encore sujet à l'expérimentation (Kunt, 2000). Ainsi, avant d'être en mesure d'estimer l'utilité d'une caractéristique plusieurs expérimentations sont généralement requises.

3.2. L'application d'une méthode de classification

La recherche d'information musicale est un domaine de recherche multidisciplinaire dédié au traitement automatisé de la musique. Ce domaine exploite des notions associées aux sciences cognitives, à la musicologie, à l'apprentissage automatisé, au traitement du signal, etc. La recherche d'information musicale est souvent représentée par l'acronyme MIR qui fait référence à l'expression anglaise « Music Information Retrieval ». Les travaux réalisés dans ce domaine ont mené au développement de plusieurs systèmes capables de classer automatiquement des musiques. La plupart des approches proposées utilisent les mêmes outils de classification que ceux exploités dans d'autres domaines comme le traitement automatisé des langues naturelles par exemple (Rompré et al., 2008). Sur le plan informatique, il existe plusieurs mécanismes pour automatiser le processus de classification. Les algorithmes de classification axés sur le voisinage comme l'algorithme Knn, le clustering, les réseaux de neurones et les SVM comptent parmi les outils informatiques couramment utilisés pour automatiser la classification des données.

3.2.1. La classification selon le voisinage

La classification basée sur le concept de voisinage est simple et intuitive. Supposons un ensemble d'entités dont la classe de similarité est connue. Ces entités forment un ensemble de référence utilisé pour guider la classification. Lorsqu'une nouvelle entité est introduite, la distance entre cette entité et celles dont la classe de similarité est connue est calculée. La nouvelle entité est assignée à la même classe que celle de l'entité dont elle est la plus similaire. La figure 3.3 illustre cette mécanique.



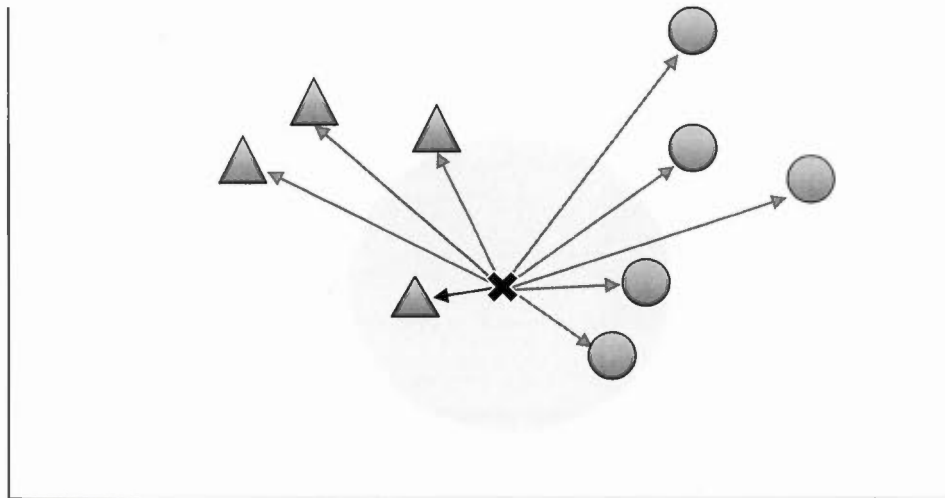
3.3. Classification basée sur le plus proche voisin

Comme la distance la plus courte mène à la classe représentée par les triangles alors la nouvelle entité est considérée comme appartenant à cette classe.

Différentes mesures sont utilisées pour calculer la distance entre deux entités. La distance Euclidienne est l'une des distances les plus couramment utilisées (Pinquier, 2004).

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \text{ où } m \text{ correspond au nombre de dimensions} \quad (3.3)$$

La technique des Knn est une variante de la classification basée sur le voisinage. Plutôt que de considérer uniquement l'entité la plus similaire, les k plus proches voisins sont considérés.



3.4. Classification basée sur les k plus proches voisins avec $k = 3$

Comme la classe de deux des trois plus proches voisins est la classe représentée par les cercles alors la nouvelle entité sera considérée comme appartenant à la classe représentée par les cercles.

Supposons que k soit fixé à 3 alors les 3 entités les plus proches de l'entité à classer sont utilisées pour déterminer à quelle classe de similarité cette dernière appartient. Dans ce cas, la classe assignée est la classe majoritaire parmi les 3 entités considérées. Étant donné sa simplicité et son niveau d'efficacité acceptable, Knn est fréquemment utilisé à titre de référence dans les études comparatives des méthodes de classification (Anderson, 1995).

3.2.2. Le clustering

Le clustering consiste à partitionner des données de manière à ce que celles qui partagent des similarités soient regroupées. Soit D un ensemble de données de taille n et k un nombre de partitions souhaitées, le problème auquel doit faire face les méthodes de clustering se résume à partitionner les données de l'ensemble D au sein des k clusters sans que la définition des clusters ne soit préalablement connue. Dans la littérature, les termes cluster et partition sont synonymes. Considérons $C = \{c_1, c_2, c_3, \dots, c_k\}$ une collection de k groupes composées d'éléments de D alors C est un clustering de D si et seulement si (Bouveryon, 2006):

- i. $c_i = \emptyset$ pour $i = 1, \dots, k$
- ii. $\bigcup_{i=1}^k c_i = D$
- iii. $c_i \cap c_j = \emptyset$ pour tout $i \neq j$

Plusieurs modèles omettent les contraintes (ii) et (iii) de manière à, d'une part, exclure les données jugées impropres à la classification et, d'autre part, autoriser l'appartenance à plus d'une classe de similarité. Ainsi, on nomme classification tout modèle respectant au minimum la contrainte (i) et clustering tout modèle

respectant à la fois les contraintes (i), (ii) et (iii). Le clustering est donc une forme spécialisée de classification. Le principe sur lequel reposent la majorité des algorithmes de clustering est qu'il existe un point central nommé centroïde qui est en mesure de généraliser la forme des membres d'un cluster. L'assignation d'une entité à un cluster est réalisée en fonction de ce point central. Un centroïde est calculé pour chacun des k clusters. Les données sont simplement assignées au cluster dont le centroïde coïncide le plus avec leurs valeurs. Soit $C = \{c_1, c_2, c_3, \dots, c_k\}$, le centroïde d'un cluster c_i , noté μ_i , peut être calculé à partir de l'équation 3.4 où n_i , e_j , et u_j correspondent respectivement à la taille de c_i , au $j^{\text{ème}}$ membre de c_i et aux valeurs de x_j .

$$\mu_i = \frac{1}{n_i} \sum_{e_j \in c_i} u_j \quad (3.4)$$

L'algorithme K-Means est l'un des algorithmes de clustering les plus communs (McQueen, 1967). L'assignation des données aux clusters est effectuée à partir d'un processus itératif. Au départ, les k centroïdes sont initialisés arbitrairement avec une valeur contenue dans l'intervalle de la valeur des unifs. Lors de chaque itération, les entités sont d'abord associées au centroïde auquel elle se rapproche le plus. L'erreur au carré est la mesure la plus communément utilisée comme mesure discriminante (équation 3.5). L'algorithme cherche à minimiser la somme de ces erreurs (équation 3.6).

$$e_i \leftarrow \min_{i=1}^k (\|e_i - \mu_i\|^2) \quad (3.5)$$

$$SEC(C) = \min(\sum_{i=1}^k \sum_{e_j \in c_i} \|e_j - \mu_i\|^2) \quad (3.6)$$

Lorsque toutes les entités sont associées à un centroïde ces derniers sont mis à jour. Le processus cesse lorsque la mise à jour des centroïdes devient négligeable. L'équation 3.7 résume ce procédé où μ_i^t correspond au centroïde du cluster c_i à l'itération t et $\mu_i^{(t-1)}$ au centroïde de ce même cluster à l'itération $t-1$. (Zaki et Meira, 2014).

$$K-Means = \sum_{i=1}^k \|\mu_i^t - \mu_i^{(t-1)}\|^2 \leq s \text{ où } s > 0 \quad (3.7)$$

Sélectionner k valeurs comme centroïde initial

Faire

Former k cluster en assignant les entités au centroïde le plus prêt

Recalculer la valeur des centroïdes

Tant que les centroïdes continuent à changer

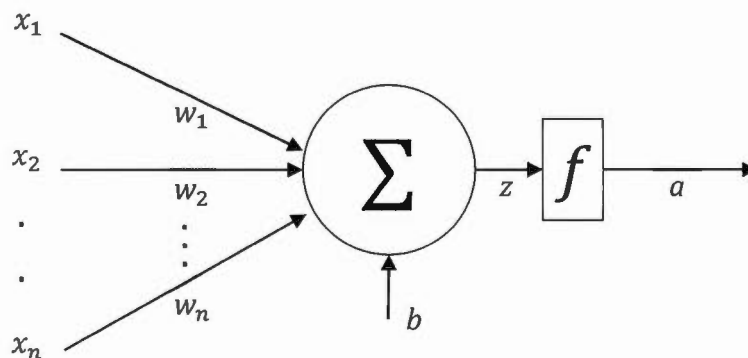
Exemple 3.1: Pseudo-code de l'algorithme K-Mean

L'algorithme K-Means fait partie de la famille des algorithmes à apprentissage non supervisé, car aucune connaissance préalable des clusters à définir n'est requise. Les clusters sont créés à suite à une évaluation des données elles-mêmes. La plus grande limitation de cette approche est que le nombre de classes doit être préalablement connu.

3.2.3. Les réseaux de neurones

Grâce à leur aptitude à capturer des propriétés structurantes à même les données brutes, les réseaux de neurones artificiels représentent des outils intéressants pour assister la création de classifications de données numériques. Ils ont la capacité à capturer des régularités au sein des données et à tolérer des entrées dont la qualité peut varier (Anderson, 1995).

Les réseaux de neurones artificiels s'inspirent des réseaux de neurones biologiques tels que ceux qui composent le cerveau de l'humain. Un réseau de neurones artificiels est un système composé d'une multitude d'unités de calcul qui sont des représentations simplifiées du neurone biologique (figure 3.5).



3.5. Schéma d'un neurone artificiel

Illustration inspirée des notes de Parizeau (Parizeau, 2006).

Les neurones artificiels sont reliés entre eux par des connexions pondérées. C'est le poids de ces connexions qui dicte le comportement du réseau. Typiquement, le neurone artificiel est constitué de deux composantes : un intégrateur qui effectue la somme pondérée des entrées et une fonction d'activation qui détermine une sortie selon la somme calculée. Il existe toutefois d'autres variantes qui cherchent à modéliser davantage la complexité du neurone biologique (Anderson, 1995). La figure 3.5 illustre un neurone artificiel typique.

Bien qu'il existe de nombreux modèles, la plupart des neurones réagissent en fonction d'un seuil d'activation qui détermine leur tolérance (Haykin, 1994). Lorsque la somme pondérée calculée par l'intégrateur est supérieure au seuil d'activation le signal envoyé vers la fonction d'activation est positif. À l'inverse, lorsque la somme pondérée est inférieure au seuil d'activation alors le signal envoyé est négatif. Typiquement, la sortie de l'intégrateur est donnée par l'équation 3.8 où a représente la sortie, n le nombre d'entrées (connexions), w_i le poids de la connexion i , x_i l'entrée i et b le seuil d'activation du neurone artificiel.

$$a = \sum_{i=1}^n w_i x_i - b \quad (3.8)$$

La sortie du neurone est soumise à une fonction d'activation. Cette fonction peut prendre plusieurs formes. La fonction sigmoïde a longtemps été l'une des plus utilisées (équation 3.9).

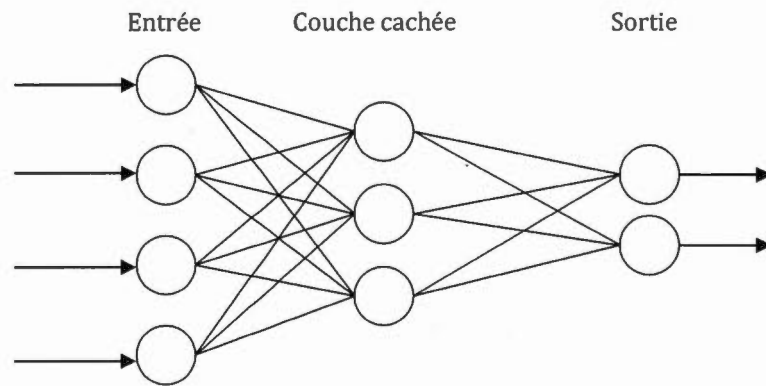
$$f(z) = \frac{1}{1 + e^{-z}} \quad (3.9)$$

La recherche effectuée ces dernières années sur les réseaux de neurones profonds a encouragé l'utilisation de la fonction d'activation à redressement¹¹. En effet, Glorot et ses collègues ont démontré que l'usage de cette fonction d'activation permettait d'accélérer l'apprentissage des réseaux (Glorot et al., 2011). La fonction d'activation à redressement est donnée par l'équation 3.10 :

$$f(z) = \max(0, z) \quad (3.10)$$

Le poids d'une connexion représente l'influence de la connexion sur l'intégrateur et éventuellement sur la sortie. Un poids positif influence positivement l'intégrateur tandis qu'un poids négatif l'influence négativement. Par conséquent, c'est en modifiant le poids des connexions entre les neurones artificiels que les connaissances sont modélisées. Bien qu'il existe différentes topologies, les réseaux de neurones sont généralement composés d'une couche d'entrée, d'une ou plusieurs couches cachées et d'une couche de sortie. Plus le réseau possède de couches cachées, plus il est considéré profond.

¹¹ Les neurones qui utilisent cette fonction sont communément appelés ReLU (*Rectifier Linear Unit*)



3.6. Exemple d'un réseau de neurones

Chacun des neurones qui composent le réseau est une unité de calcul comme celle illustrée à la figure 3.5.

L'inscription de connaissances à l'intérieur du réseau est réalisée suite à un processus d'apprentissage (Haykin, 1994). Ce processus permet au réseau d'améliorer ses performances. Le type d'apprentissage correspond à la méthode utilisée pour adapter le poids des connexions du réseau et ou la valeur des seuils d'activation des neurones. Souvent, plusieurs expérimentations sont nécessaires pour entraîner favorablement un réseau. Il existe deux grandes familles d'apprentissages soit l'apprentissage supervisé et non supervisé.

3.2.3.1. L'apprentissage non supervisé

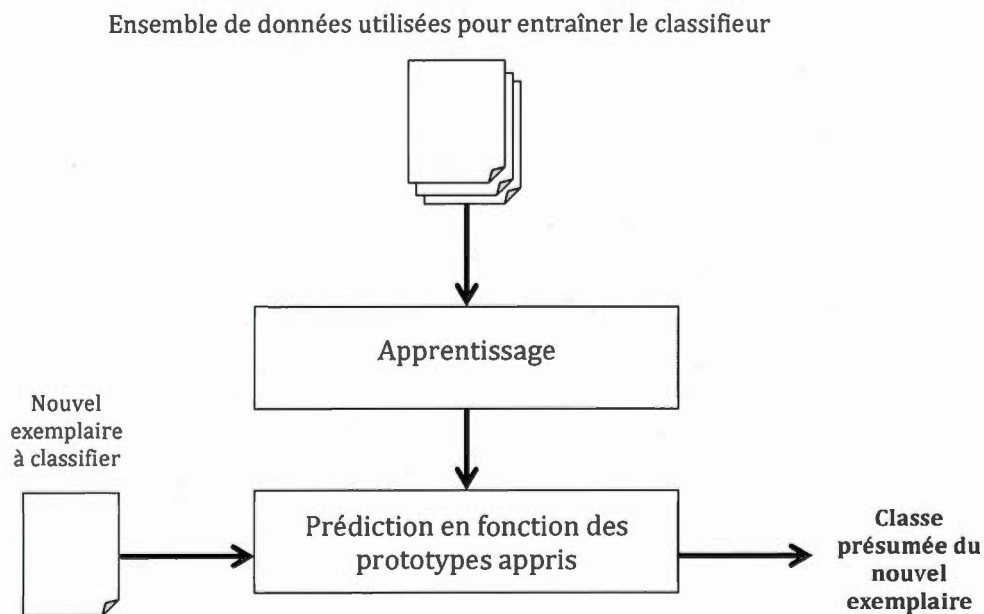
L'apprentissage non supervisé consiste à laisser le réseau adapter les poids de ses connexions sans que sa sortie soit évaluée. Les stratégies employées pour ce faire s'inspirent de principes associés à la neurobiologie telle que l'apprentissage hebbien. L'apprentissage hebbien consiste à renforcer le poids de la connexion entre les neurones qui sont actifs simultanément et à atténuer le poids des connexions des neurones qui sont activés de manière asynchrone. La règle de Hebb qui traduit ce principe permet d'ajuster le poids des connexions (Anderson, 1995). Cette règle est donnée par l'équation 3.11 où η représente un facteur d'apprentissage prédéterminé, x_i l'activité du neurone pré synaptique et y_i l'activité du neurone post synaptique.

$$\Delta_{ij} = \eta x_i y_i \quad (3.11)$$

L'apprentissage se fait localement. La mise à jour du poids de la connexion entre un neurone k et un neurone l dépend uniquement de l'activité de ces neurones et nullement de la sortie. Cette caractéristique permet aux réseaux de capturer des régularités à même les données brutes sans connaître la nature de ces données. L'un des problèmes de l'apprentissage hebbien est que les poids peuvent croître constamment si la même paire d'entrée-sortie est présentée continuellement. Pour corriger cette problématique, un facteur d'oubli peut être ajouté (Haykin, 1994). Cependant, un tel ajout nécessite que le réseau soit stimulé régulièrement sans quoi le réseau peut perdre des connaissances acquises.

3.2.3.2. L'apprentissage supervisé

L'apprentissage supervisé est un mode d'apprentissage qui implique que des données d'entrées dont la nature est connue soient disponibles afin qu'il soit possible d'évaluer la sortie du réseau en fonction de celle qui est désirée. La figure 3.7 illustre cette particularité.



3.7. Processus de classification supervisée

L'apprentissage est effectué grâce à la rétro-propagation de l'erreur de la couche de sortie vers la couche d'entrée. Le poids des connexions est ajusté en fonction de l'erreur calculée. Soit x un vecteur d'entrées, a un vecteur de sorties

observées et y un vecteur des sorties désirées, alors l'erreur de sortie d'un neurone i est exprimée par l'équation 3.12 :

$$e_i = y_i - a_i \quad (3.12)$$

L'objectif de la rétro-propagation de l'erreur est d'ajuster les combinaisons de poids et de seuils d'activation afin de permettre aux réseaux de produire la sortie désirée. Pour ce faire, une fonction de calcul du coût est exploitée. Typiquement, la somme des erreurs quadratique est utilisée (équation 3.13).

$$C = \frac{1}{2} \sum_j (y_j - a_j)^2 \quad (3.13)$$

L'ajustement est effectué à partir de la descente du gradient. L'application de cette opération d'optimisation a comme objectif de trouver les poids qui minimisent la fonction de coût. Les poids sont modifiés dans le sens opposé du gradient $\frac{\partial C}{\partial w}$ de l'erreur observée. La variation des poids est donnée par l'équation 3.14 où η est un scalaire en 0 et 1 qui représente le taux d'apprentissage.

$$\Delta w = \eta \frac{\partial C}{\partial w} \quad (3.14)$$

À l'aide de la règle de chaînage des dérivées partielles il est possible de fixer la valeur du gradient, dans un premier temps, pour la couche de sortie et par la suite pour les couches cachées du réseau¹². Ainsi, les poids sont ajustés en rétro-propageant l'erreur de la sortie vers l'entrée.

3.2.3.3. Les réseaux de neurones profonds

L'intérêt pour le développement d'architectures capables de traiter divers niveaux d'abstraction de données a influencé l'évolution des réseaux de neurones. Les réseaux profonds forment des modèles computationnels composés de plusieurs couches dont le rôle est de manipuler des données dont le niveau d'abstraction varie. La représentation de l'information propagée dans les réseaux profonds se transforme d'une couche à l'autre en commençant par les données brutes en entrée vers une forme de représentations plus significative et généralisée. Par exemple, dans un contexte de reconnaissance de forme, la valeur des pixels qui composent une image peut être soumise à l'entrée du réseau. L'information capturée par la couche d'entrée peut alors être propagée et être utilisée par la couche suivante pour inscrire la présence de lignes ou de courbes. Les lignes et les courbes inscrites peuvent alors être à leur tour utilisées par une couche suivante pour déduire la présence des formes géométriques dans l'image. Les informations capturées ne sont pas modélisées par l'humain mais plutôt apprises à l'aide d'algorithmes d'apprentissage basés sur la rétro-propagation de l'erreur (LeCun et al., 2015).

Avant 2006, les résultats obtenus à l'aide de réseaux de neurones profonds étaient plutôt mitigés. La difficulté à entraîner les réseaux profonds était l'une

¹² Pour plus de détails consultez : <http://neuralnetworksanddeeplearning.com/>

des raisons évoquées pour expliquer ces résultats. L'initiation aléatoire des poids au démarrage du processus d'apprentissage a été identifiée comme une des sources du problème (Hinton et al., 2006). Afin de favoriser l'apprentissage des réseaux profonds, Hinton et ses collègues ont introduit une procédure de pré-entraînement. Cette procédure consiste à pré-entraîner chaque couche du réseau à l'aide d'une phase d'apprentissage non supervisé. L'apprentissage se fait une couche à la fois. Les représentations apprises par les couches inférieures sont utilisées pour entraîner les couches supérieures. Finalement, une étape d'apprentissage supervisé est effectuée pour raffiner les performances du réseau (Benjio et al., 2007).

Le choix de la fonction sigmoïde comme fonction d'activation a également été identifié comme étant un obstacle pour l'entraînement des réseaux profonds. Au début de la présente décennie, il a été démontré que l'utilisation de la fonction de redressement comme fonction d'activation permet d'éviter d'avoir recours à l'étape de pré-entraînement non supervisé (Glorot et al., 2011). Les avancées technologiques ont également contribué au développement des réseaux profonds. En effet, l'utilisation de processeurs graphiques optimisés pour le calcul matriciel permet de réduire substantiellement le temps nécessaire pour entraîner les réseaux profonds (Raina et al., 2009).

Les récentes percées dans le domaine de l'apprentissage supervisé ont entraîné la création de nombreuses architectures profondes telles que les réseaux DBN¹³ et les DCN¹⁴. Ces architectures jouissent actuellement d'une certaine popularité due à leur capacité à reconnaître des formes complexes. Pour cette raison, ils ont été utilisés avec succès pour mettre en œuvre des systèmes de classification

¹³ *Deep Believe Network*

¹⁴ *Deep Convolutional Network*

d'image (Le et al., 2012) et de reconnaissance de la parole (Hinton et al., 2012). Les réseaux de neurones profonds composés quelques dizaines de couches cachées peuvent à la fois être sensibles à certains détails permettant de différencier certains éléments ressemblants et à la fois être insensibles à des variations importantes non pertinentes (LeCunn et al., 2015).

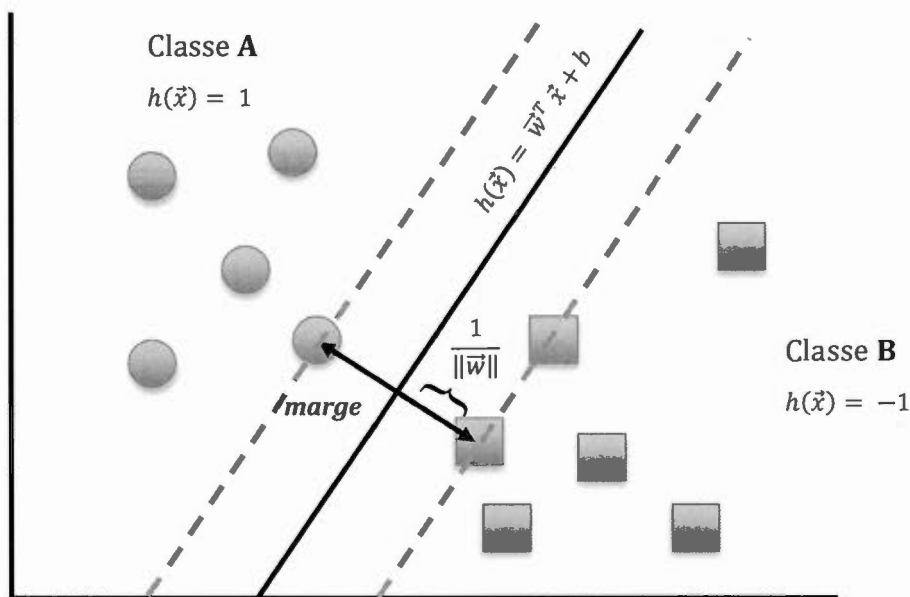
Dans un contexte de classification, les réseaux de neurones sont entraînés pour reconnaître les patrons qui caractérisent les classes. Un des problèmes des réseaux de neurones est que leur performance dépend de la qualité de l'échantillonnage utilisé pour les entraîner (Sutton et Barto, 1998). Un échantillonnage représentatif permet au réseau de bien performer lorsque de nouvelles entrées lui sont présentées. Toutefois, il n'existe aucune garantie que cet échantillonnage existe. De plus, la portée des réseaux de neurones se limite la plupart du temps au domaine pour lequel ils sont entraînés. Ces problèmes demeurent même pour les réseaux de neurones profonds.

Le succès des modèles computationnels qui exploitent des réseaux de neurones profonds illustre l'importance des niveaux d'abstraction dans le traitement de l'information. Toutefois, les réseaux de neurones profonds sont composés de plusieurs millions de connexions ce qui rend difficile l'analyse de leur structure interne.

3.2.4. Les SVM

Les SVM (Support Vector Machine) forment une autre méthode d'apprentissage supervisé. Le développement de cette méthode résulte des travaux de Vapnik (Vapnik, 1995). Depuis l'introduction des SVM en 1995, il existe un certain

enthousiasme par rapport à leur usage pour résoudre des problèmes liés à la classification automatisée. Le principe de base des SVM consiste à maximiser la marge qui sépare deux ensembles de données (figure 3.8).



3.8. Les SVM

Les SVM cherchent à maximiser la marge entre deux ensembles.

Dans le cas où l'ensemble de données est linéairement séparable, la droite qui subdivise cet ensemble en deux est donnée par l'équation 3.15 où w donne l'orientation de la droite et b sa position. Chaque demi-espace ainsi créé correspond à une classe. Par conséquent, cette droite équivaut à la frontière entre les deux classes.

$$h(\vec{x}) = \vec{w}^T \vec{x} + b, \text{ tel que } h(\vec{x}) = \begin{cases} 1 & \text{si } \vec{x} \in A \\ -1 & \text{si } \vec{x} \in B \end{cases} \quad (3.15)$$

Il existe une infinité de frontières admissibles. La frontière optimale est considérée comme étant l'hyperplan séparateur situé à mi-chemin entre les éléments les plus proches de chacune des classes. La distance entre la frontière et l'élément qui est le plus proche de cette frontière est donnée par l'équation 3.16. Ainsi la marge optimale, donnée par l'équation 3.17, équivaut à deux fois cette distance.

$$d = \frac{1}{\|\vec{w}\|} \quad (3.16)$$

$$\text{marge} = \frac{2}{\|\vec{w}\|} \quad (3.17)$$

Par conséquent, en minimisant \vec{w} la marge peut être maximisée. Cette opération est un problème d'optimisation qui peut être résolu à l'aide des conditions de Karush, Kuhn et Tucker (Zaki et Meira, 2014) et de la méthode des multiplicateurs de Lagrange (Zaki et Meira, 2014). Afin de pouvoir traiter les ensembles non linéairement séparables, les SVM exploitent la notion de noyau non linéaire. Le rôle d'un noyau est de transposer les éléments dans une dimension supérieure dans laquelle il est probable qu'ils puissent être séparés. La figure 3.9 illustre cette démarche.



3.9. Transposition dans une autre dimension

La fonction des noyaux consiste à transposer les éléments dans une autre dimension supérieure dans laquelle ils sont susceptibles d'être séparables

3.2.5. Discussion au sujet du MIR et de la classification

Depuis le début des années 2000, plusieurs chercheurs se sont intéressés au problème de classification de données sonores. Le développement de ces travaux porte d'une part sur l'extraction de caractéristiques pour représenter les propriétés de l'onde sonore et d'autre part sur l'évaluation de nouvelles stratégies de classification des données sonores. Même s'il existe une certaine controverse au sujet l'usage du genre comme critère de classification, la recherche dédiée à la classification de musique par rapport au genre musical est un des axes de recherche les plus actifs du MIR.

Les travaux de Tzanetakis sont considérés comme des travaux précurseurs dans ce domaine (Tzanetakis, 2002 ; Tzanetakis et Cook, 2002). Tzanetakis a évalué différentes mesures jumelées à différents classifieurs. Les meilleures classifications produites lors de ces expérimentations ont été jugées cohérentes

à 61%. Suite à l'analyse de ses résultats, Tzanetakis est venu à la conclusion que le rock est le genre le plus difficile à classifier et ce peu importe les mesures et le classifieur utilisés. Selon Tzanetakis, ce phénomène est lié au fait que le rock représente un genre musical qui englobe plusieurs formes de musiques (pop-rock, hard rock, etc.) et qui entrecoupe parfois d'autres genres musicaux (métal, rock'n roll, etc.).

Les travaux de McKay et Fujinaga sont également cités à de nombreuses reprises dans la littérature (McKay et Fujinaga, 2004). Dans le cadre de leurs travaux, ils ont exploré une approche qui mixte Knn et des réseaux de neurones supervisés (McKay et Fujinaga, 2004). Le meilleur taux de précision obtenu lors de leurs expérimentations est de 98 % lorsque plusieurs descripteurs sont utilisés pour représenter le contenu des fichiers sonores. Cependant, ce score est rapporté uniquement lorsque le nombre de genres considérés est limité à 3 (classique, jazz, populaire). Lorsque leurs expérimentations intègrent 9 genres, le score obtenu chute de 8 % pour se situer à 90% ce qui demeure néanmoins intéressant. L'une des modalités de la solution proposée par McKay et Fujinaga est qu'elle opère sur des fichiers de format MIDI. La particularité de ce format de fichier est qu'il contient une description détaillée de la mélodie. Ainsi, un fichier MIDI fournit explicitement le ton, l'intensité, et la durée des notes. Cette particularité rend plus aisée l'extraction des propriétés qui caractérisent la musique. Cependant, la majorité des fichiers sonores contenus dans les bases de données ne sont pas de ce format ce qui limite la portée des travaux de McKay et Fujinaga.

Deshpande et ses collègues ont expérimenté l'usage de Knn et de K-means pour classifier 157 pièces musicales au sein de 3 classes représentant le classique, le rock et le jazz (Deshpande et al., 2001). Leurs expérimentations suggèrent que la musique classique s'adapte davantage au processus de classification automatisée que la musique rock et jazz. En fait, aux dires des auteurs eux-mêmes, leur modèle est en mesure de classifier adéquatement uniquement la musique classique, et ce avec un niveau de précision de 73%. L'écart entre les résultats de Deshpande et ceux de McKay et Fujinaga illustre la difficulté associée à l'automatisation de la classification de musique à partir du traitement de l'onde sonore.

Ces dernières années, l'usage des SVM a gagné en popularité. Signe de cet intérêt, 11 des 12 soumissions au concours MIREX 2014¹⁵ (*Music Information Retrieval Evaluation eXchange*) consacré à la classification de musique par genre faisaient usage des SVM. MIREX est un concours organisé dans le cadre de la conférence ISMIR¹⁶ (*International Society for Music Information Retrieval*). L'objectif de ce concours est de comparer l'efficacité de nouveaux modèles de classification. Les résultats du concours MIREX 2014 sont donnés au tableau 3.4.

¹⁵ http://www.music-ir.org/nema_out/mirex2014/results/act/mixed_report/

¹⁶ <http://www.ismir.net/>

Genre	BK1	BK2	BK3	BK4	QK2	SS6	SSKS1	WJ2	ZL2	ZL3	ZL4	Moyenne
<i>baroque</i>	82,14	76,29	63,71	69,71	74,71	80,86	80,86	89,32	77,86	78,14	83,71	71,44
<i>blues</i>	75,43	70,43	62,86	66,71	70,14	79,71	79,71	87,27	70,71	68,71	72,86	67,05
<i>classical</i>	74,86	70,14	64,00	62,29	63,71	72,43	72,43	76,64	76,86	71,86	69,86	64,59
<i>country</i>	79,14	74,86	68,14	69,43	74,29	78,29	78,29	87,00	74,71	72,86	66,43	68,62
<i>edance</i>	79,14	74,29	71,43	77,71	62,71	78,71	79,00	85,26	0,00	0,00	74,86	56,93
<i>jazz</i>	75,29	68,00	58,29	60,43	65,43	82,71	82,71	89,70	69,86	68,43	72,00	66,07
<i>metal</i>	72,29	74,71	72,43	76,00	75,86	72,57	72,57	83,95	70,43	68,57	75,71	67,92
<i>rap hiphap</i>	90,57	89,14	87,71	85,43	89,00	88,57	88,57	91,71	0,00	0,00	81,14	65,99
<i>rockroll</i>	56,00	50,14	40,57	38,14	45,14	55,71	55,57	70,44	0,00	0,00	30,43	36,85
<i>romantic</i>	62,86	54,00	48,14	56,86	52,43	65,29	65,29	74,18	63,57	62,43	46,43	54,29
Moyenne	74,77	70,20	63,73	66,27	67,34	75,49	75,50	83,55	50,40	49,10	67,34	61,97

- BK1** Seung Ryoel Baek et Moo Young Kim (SVM)
BK2 Seung Ryoel Baek et Moo Young Kim (SVM)
BK3 Seung Ryoel Baek et Moo Young Kim (SVM)
BK4 Seung Ryoel Baek et Moo Young Kim (SVM)
QK2 Qiuqiang Kong (Réseau de neurones)
SS6 Klaus Seyerlehner et Markus Schedl (SVM)
SSKS1 Klaus Seyerlehner, Markus Schedl, Peter Knees et Reinhard Sonnleitner (SVM)
WJ2 Ming-Ju Wu et Jyh-Shing Roger Jang (SVM)
WJ2 Ming-Ju Wu et Jyh-Shing Roger Jang (SVM)
ZL2 Xinquan Zhou, Guang Li (SVM)
ZL3 Xinquan Zhou, Guang Li (SVM)
ZL4 Xinquan Zhou, Guang Li (SVM)

3.4. Résultat du concours MIREX 2014

3.3. L'interprétation des résultats

La classification est une tâche qui organise les données de manière à ce que celles qui partagent des propriétés soient regroupées dans des classes de similarités. L'organisation qui en découle peut être utilisée pour orienter l'exploration des données. Comme présenté précédemment, les propriétés considérées lors du processus de classification ne sont pas directement accessibles. Des caractéristiques extraites des données brutes doivent être utilisées pour représenter les propriétés convoitées. Le choix d'une caractéristique aux dépens d'une autre revient en quelque sorte à prendre position sur la nature des résultats qui seront produits, car le choix des caractéristiques influence le comportement du classifieur. Dans le même ordre d'idée, le choix du classifieur lui-même influence la classification résultante. Les figures 3.3 et 3.4 illustrent ce phénomène. Pour un même ensemble de données, les résultats ne sont pas les mêmes. À ce jour, il n'existe pas de modèle de classification optimal. La valeur des résultats générés dépend à la fois des caractéristiques, du classifieur et de l'ensemble de données considérés.

Les meilleurs résultats présentés dans la littérature ont été obtenus lorsque les données considérées réfèrent à peu de classes. Autrement dit, les modèles de classification automatisée sont en mesure de bien performer lorsque la complexité du schéma de classification abordé demeure relativement simple. Des concours tels que le concours MIREX sont organisés pour mesurer les progrès réalisés dans le développement des méthodes de classification automatisée de données musicales. L'une des problématiques rencontrées est que les méthodes développées sont fréquemment spécialisées dans le traitement d'un ensemble de données prédéterminé. La musique est complexe et diversifiée. Peu importe la méthode employée pour classer les données

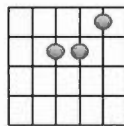
sonores, la complexité, la diversité et le caractère multidimensionnel de la musique entraînent l'introduction de bruit dans les classes de similarité. Le bruit introduit lors du processus de classification et le nombre potentiellement élevé de classes peuvent alors donner l'impression que les classes de similarités générées sont peu ou pas significatives. Cependant, une analyse des classes de similarité peut faire ressortir des propriétés intéressantes et inattendues. L'analyse de ces classes est toutefois un exercice non trivial.

CHAPITRE IV

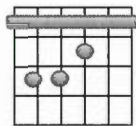
LES RÈGLES D'ASSOCIATION

La recherche effectuée depuis le début des années 1990 dans le domaine du data mining a mené au développement de plusieurs outils capables de dégager des connaissances cachées à l'intérieur de vastes ensembles de données. L'extraction de règles d'association est une technique qui figure parmi ces outils. Cette technique consiste sommairement à trouver des règles qui prédisent l'occurrence d'une donnée en fonction de l'occurrence d'autres données (Tan et al., 2006). L'intérêt porté à l'extraction de règles d'association découle grandement des travaux d'Agrawal sur la recherche de connaissances enfouies dans des bases de données transactionnelles (Agrawal et al., 1993 ; Agrawal et Strikant, 1994). Originellement, la méthode proposée par Agrawal était destinée à illustrer des relations entre des items qui cooccurrent dans des transactions commerciales. Par exemple, les clients qui achètent les items x et y achètent également l'item z . Graduellement, le concept de transaction a été redéfini selon les domaines abordés. Par exemple, une transaction peut représenter un historique de navigation sur Internet (Likari et al., 2015). Ainsi, l'expression « *les clients qui achètent les items x et y achètent également l'item z* » devient « *les utilisateurs qui visitent les sites web x et y visitent également le site z* ». Les règles d'associations s'appliquent à divers domaines dans la mesure où des transactions peuvent y être définies.

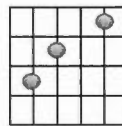
Accords :



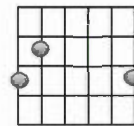
Am



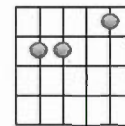
F



C



G



CB

Intro :

Am F C (x 2)

Couplet 1 :

Am G C
 Quand on verra dans le miroir nos faces ridées pleines d'histoires
 Am G C
 Quand on en aura moins devant qu'on en a maintenant
 Am G C
 Quand on aura enfin du temps et qu'on vivra tranquillement
 Am G C
 Quand la maison sera payée et qui restera plus rien qu'à s'aimer . .

Refrain :

Am G F C CB
 On va s'aimer encore, au travers des doutes, des travers de la route et de plus en plus fort on va s'aimer encore

Am G F C CB
 Au travers des bons coups, au travers des déboires, à la vie, à la mort on va s'aimer encore

Transition :

Am F C

Exemple 4.1 : Extrait des paroles et accords de la chanson « On va s'aimer encore » de Vincent Vallières. Le texte en bleu représente les accords de guitare.

Dans la littérature, le terme transaction demeure couramment employé comme synonyme d'un enregistrement dans une base de données (Selvi et Tamilarasi, 2009 ; Tan et al., 2006). Toutefois, cette utilisation est limitative. Il convient donc de préciser ce que représente une transaction de manière plus générale. Globalement, une transaction désigne un regroupement de données descriptives. Cette définition plus large autorise diverses formes d'analyses. Par exemple, une analyse hypothétique d'une pièce musicale pourrait être effectuée en utilisant les successions d'accords de cette pièce en guise de transactions. L'exemple 4.1 donne un extrait des paroles et des accords de guitare de la chanson « On va

s'aimer encore » de l'artiste Vincent Vallières. Basé sur cet exemple, les successions d'accords « Am, F, C », « Am, G, C » et « Am, G, F, C/B » représentent des transactions admissibles.

4.1. Définition des règles d'association

Une règle d'association notée $X \rightarrow Y$ traduit une relation de cooccurrence entre des données qui apparaissent fréquemment ensemble. L'objectif du processus d'extraction des règles d'association est de faire ressortir des relations significatives cachées à l'intérieur de transactions. Le processus qui mène à la découverte des règles d'association est assujetti à deux problèmes importants. Le premier est lié à l'effort computationnel requis pour exécuter la tâche. En effet, l'extraction de règles d'association exige plusieurs calculs et la somme de ces calculs croît en fonction de la quantité de données à traiter. Comme le processus d'extraction des règles d'association est destiné à être appliqué à de vastes ensembles de données diverses stratégies doivent être employées afin de réduire les calculs nécessaires. Le deuxième problème est lié à l'utilité de certaines règles d'association. Plusieurs règles peuvent être générées à partir d'un vaste ensemble de données. Parmi ces règles, certaines sont uniquement le résultat du hasard et par conséquent elles sont habituellement sans intérêt. C'est pourquoi des mesures doivent être utilisées pour estimer l'utilité des règles générées. Le concept de règles d'association est étroitement lié aux notions d'itemset, de support et de confiance. Ces notions introduites par Agrawal sont exploitées par la majorité des solutions proposées pour répondre aux deux grandes problématiques citées précédemment. Cette section présente ces notions fondamentales telles que proposées par Agrawal.

Soit T un ensemble composé de n transactions tel que $T = \{t_1, t_2, t_3, \dots, t_n\}$, les éléments qui composent les transactions $t_i \in T$ sont appelés des *items*. Cette appellation tire son origine des travaux d'Agrawal en référence aux items qui cooccurrent dans des transactions commerciales. Un item est une donnée dont la nature dépend du domaine abordé. Dans l'exemple 4.1 les accords de guitare $\{A_M\}$, $\{F\}$, $\{C\}$, $\{G\}$ et $\{C_B\}$ constituent les items. Le tableau 4.1 donne la répartition de ces items dans les transactions.

Transactions	Items				
	A_M	F	C	G	C_B
1	x	x	x		
2	x	x	x		
3	x		x	x	
4	x		x	x	
5	x		x	x	
6	x		x	x	
7	x	x	x	x	x
8	x	x	x	x	x
9	x	x	x		

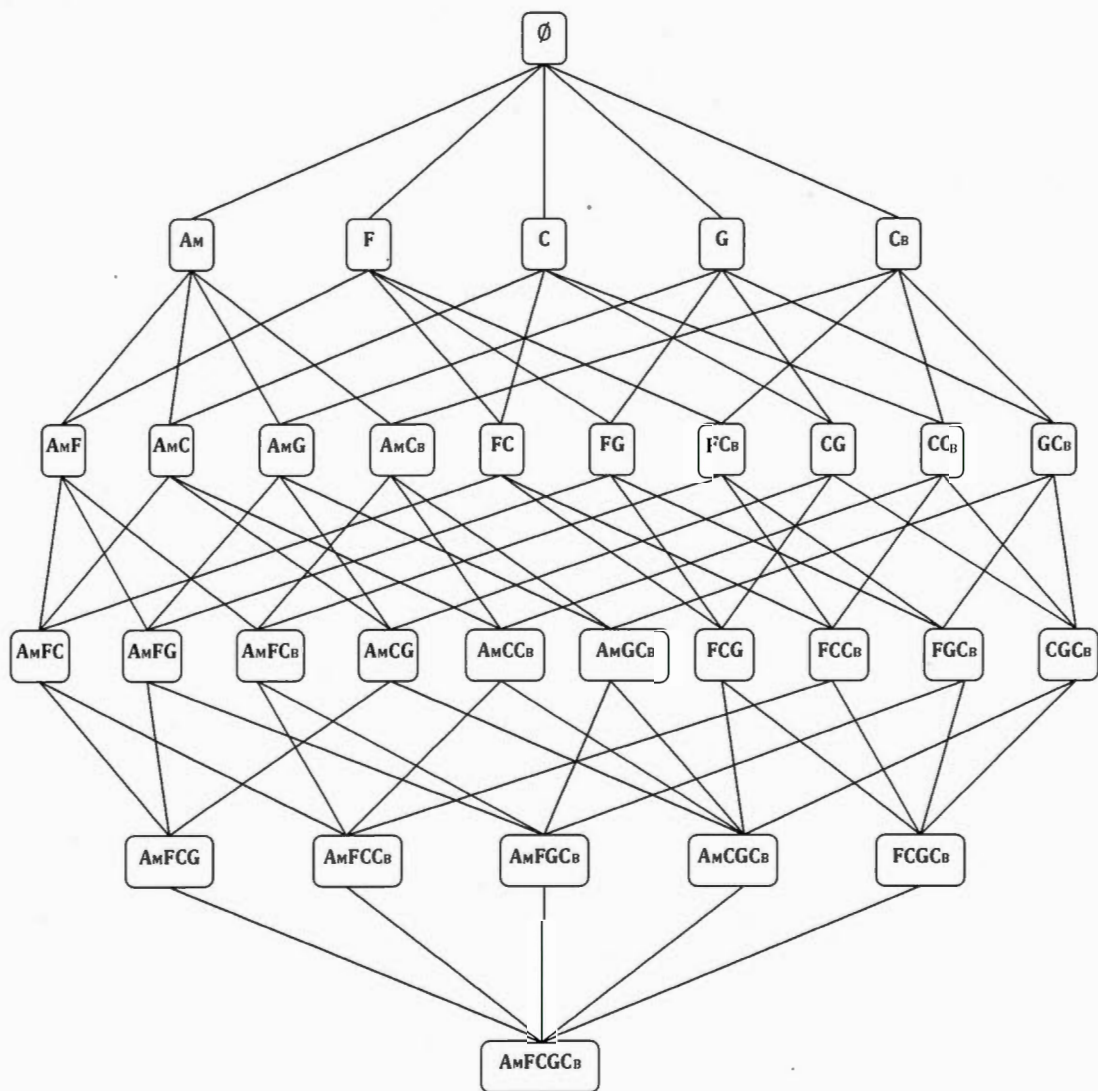
4.1. Répartition des items dans les transactions

Répartition des itemsets dans les transactions de l'exemple 4.1

4.1.1. Itemset

Soit $I = \{i_1, i_2, i_3, \dots, i_d\}$ un ensemble composé de d items distincts, chaque sous-ensemble qu'il est possible de générer à partir des items $i_i \in T$ est appelé un itemset. Pour un ensemble I de taille d , le nombre d'itemsets possibles est 2^d . Le nombre d'itemsets potentiels est donc exponentiel en fonction de la taille de I .

La figure 4.1 illustre le diagramme de Hasse de tous les itemsets qu'il est possible de générer à partir des cinq items de l'exemple 4.1.



4.1. Diagramme de Hasse

Le diagramme de Hasse de la figure 4.1 représente les itemsets qu'il est possible de créer à partir des items $\{A_M\}$, $\{F\}$, $\{C\}$, $\{G\}$ et $\{C_B\}$.

La notion d'itemset est importante, car c'est à partir de ces sous-ensembles que les règles d'association sont produites. Comme l'objectif à atteindre lors du processus d'extraction des règles d'association est de découvrir des relations cachées, il n'y a pas d'indice permettant de cibler les items ou les ensembles d'items à considérer. Ainsi, l'espace de recherche équivaut à l'ensemble des itemsets possibles. Même s'il est théoriquement possible de créer 2^d itemsets à partir d'un ensemble de taille d , en pratique plusieurs combinaisons apparaissent peu ou n'apparaissent tout simplement pas dans les transactions. Du coup, ces combinaisons peuvent être ignorées. Afin de pouvoir cibler les itemsets à ignorer, Agrawal a introduit la notion de support. Le support est une mesure qui permet d'estimer l'utilité d'un itemset en fonction de sa fréquence d'apparition dans les transactions. Il existe deux variantes de cette mesure : le « support brut¹⁷ » et le « support ».

4.1.2. Support d'un itemset

Soit $I = \{i_1, i_2, i_3, \dots, i_d\}$ un ensemble de d items distincts, $T = \{t_1, t_2, t_3, \dots, t_n\}$ un ensemble de n transactions où chaque transaction t_i est composée d'items de I et X un itemset tel que $X \neq \emptyset$, le support brut d'un itemset X représente le nombre de transactions de T qui contiennent X . Le support brut d'un itemset X , noté $\sigma(X)$, est donné par l'équation 4.1 (Tan et al., 2006):

$$\sigma(X) = |\{t_i \mid X \subseteq t_i, t_i \in T\}| \quad (4.1)$$

¹⁷ En anglais l'expression « Support Count » est employée.

Le support d'un itemset est une mesure similaire au support brut. Cependant, à l'opposé du support brut, le nombre de transactions total est considéré. Ainsi, le support d'un itemset X représente le pourcentage des transactions de T qui contient X . Le support d'un itemset X , noté $S(X)$, est donné par l'équation 4.2 où n équivaut au nombre total de transactions contenues dans T .

$$S(X) = \frac{\sigma(X)}{n} \quad (4.2)$$

Le tableau 4.2 donne le support brut et le support de tous les itemsets qu'il est possible de générer à partir des items $\{A_M\}$, $\{F\}$, $\{C\}$, $\{G\}$ et $\{C_B\}$.

Itemset	$\sigma(X)$	$S(X)$	Itemset	$\sigma(X)$	$S(X)$	Itemset	$\sigma(X)$	$S(X)$	Itemset	$\sigma(X)$	$S(X)$
\emptyset	-	-	$A_M G$	6	0.67	$A_M F C$	5	0.56	$F G C_B$	2	0.22
A_M	9	1.00	$A_M C_B$	2	0.22	$A_M F G$	2	0.22	$C G C_B$	2	0.22
F	5	0.56	$F C$	5	0.56	$A_M F C_B$	2	0.22	$A_M F C G$	2	0.22
C	9	1.00	$F G$	2	0.22	$A_M C G$	6	0.67	$A_M F C C_B$	2	0.22
G	6	0.67	$F C_B$	2	0.22	$A_M C C_B$	2	0.22	$A_M F G C_B$	2	0.22
C_B	2	0.22	$C G$	6	0.67	$A_M G C_B$	2	0.22	$A_M C G C_B$	2	0.22
$A_M F$	5	0.56	$C C_B$	2	0.22	$F C G$	2	0.22	$F C G C_B$	2	0.22
$A_M C$	9	1.00	$G C_B$	2	0.22	$F C C_B$	2	0.22	$A_M F C G C_B$	2	0.22

4.2. Support brut et support

Support brut et support de tous les itemsets qu'il est possible de générer à partir des items $\{A_M\}$, $\{F\}$, $\{C\}$, $\{G\}$ et $\{C_B\}$ en fonction des transactions de l'exemple 4.1.

On appelle itemset fréquent un itemset dont le support est supérieur ou égal à un seuil préalablement fixé nommé *support minimum*. Dans le même ordre d'idée, on appelle itemset infrequent un itemset dont le support est inférieur au support minimum. Les itemsets fréquents sont réputés être les itemsets les plus intéressants. L'espace de recherche peut donc être réduit en éliminant les itemsets infrequent. Il est facile de constater que la valeur attribuée au support minimum influence considérablement la proportion des itemsets jugés fréquents par rapport à ceux jugés infrequent.

Considérons les données du tableau 4.2 et supposons que le support minimum soit fixé à 0.70 alors les itemsets fréquents sont $\{A_M\}$, $\{C\}$ et $\{A_MC\}$. Supposons maintenant que le support minimum est abaissé à 0.65 alors les itemsets deviennent $\{A_M\}$, $\{C\}$, $\{G\}$, $\{A_MC\}$, $\{A_MG\}$, $\{CG\}$, et $\{A_MCG\}$. La valeur accordée au support minimum témoigne habituellement d'un compromis entre quantité et qualité.

4.1.3. Les règles d'association

Soit X et Y deux itemsets fréquents tel que $X \cap Y = \emptyset$, une règle d'association notée $X \rightarrow Y$ traduit une relation de cooccurrence entre ces itemsets. Par convention, le premier terme est appelé l'antécédent tandis que le second est appelé le conséquent. Les expressions suivantes représentent les règles d'associations qu'il est possible de créer à partir des itemsets fréquents $\{A_M\}$, $\{C\}$, $\{G\}$, $\{A_M C\}$, $\{A_M G\}$, $\{CG\}$, et $\{A_M CG\}$: $\{A_M\} \rightarrow \{C\}$, $\{A_M\} \rightarrow \{G\}$, $\{C\} \rightarrow \{A_M\}$, $\{C\} \rightarrow \{G\}$, $\{G\} \rightarrow \{A_M\}$, $\{G\} \rightarrow \{C\}$, $\{A_M\} \rightarrow \{CG\}$, $\{G\} \rightarrow \{A_M G\}$, $\{G\} \rightarrow \{A_M C\}$, $\{A_M C\} \rightarrow \{G\}$, $\{A_M G\} \rightarrow \{C\}$ et $\{CG\} \rightarrow \{A_M\}$.

Une règle d'association est jugée de qualité selon une mesure m et un seuil préalablement fixé. Ainsi, une règle d'association $X \rightarrow Y$ est jugée de qualité si $(X \rightarrow Y) \geq \text{seuil}$. La quantité de règles générées, leur pertinence de même que leur utilité dépendent fortement des mesures et des seuils minimaux fixés. L'évaluation des mesures d'intérêt des règles d'association a fait l'objet de plusieurs travaux (Le Bras et al., 2010 ; Lenca et al., 2008 ; Geng et Hamilton, 2006; Vaillant; 2006 ; Tan et al., 2002). Le tableau 4.3 dresse une liste de quelques mesures existantes (Geng et Hamilton, 2006). Parmi toutes les mesures existantes, le support et la confiance tel que proposés par Agrawal (Agrawal et al., 1993) demeurent des plus employés.

Mesure	Formule
Support	$P(AB)$
Confiance	$P(B A)$
Couverture	$P(A)$
Prévalence	$P(B)$
Rappel	$P(A B)$
Spécificité	$P(\neg A \neg B)$
Exactitude	$P(AB) + P(\neg A\neg B)$
Intérêt	$\frac{P(B A)}{P(B)}$
Effet de levier	$P(B A) - P(A)P(B)$
Valeur ajoutée	$P(B A) - P(B)$
Risque relatif	$\frac{P(B A)}{P(B \neg A)}$
Jaccard	$\frac{P(AB)}{P(A) + P(B) + P(AB)}$
Facteur de certitude	$\frac{P(B A) - P(B)}{1 - P(B)}$
Rapport de chance	$\frac{P(AB) - P(\neg A\neg B)}{P(\neg A\neg B)P(A\neg B)}$
Yule's Q	$\frac{P(AB)P(\neg A\neg B) - P(A\neg B) - P(\neg AB)}{P(AB)P(\neg A\neg B) + P(A\neg B) - P(\neg AB)}$
Yule's Y	$\frac{\sqrt{P(AB)P(\neg A\neg B)} - \sqrt{P(A\neg B) - P(\neg AB)}}{\sqrt{P(AB)P(\neg A\neg B)} + \sqrt{P(A\neg B) - P(\neg AB)}}$
Klosgen	$\frac{\sqrt{P(AB)(P(B A) - P(B))}}{\sqrt{P(AB) \max\{P(B A) - P(B), P(A B) - P(A)\}}}$
Conviction	$\frac{P(A)P(\neg B)}{P(A\neg B)}$
Taux pondéré d'intérêt	$\left(\left(\frac{P(AB)}{P(A)P(B)} \right)^k - 1 \right) P(AB)^m$
Force collective	$\frac{P(AB) + P(\neg B \neg A)}{P(A)P(B) + P(\neg A)P(\neg B)}$
Correction de Laplace	$\frac{n(AB) + 1}{n(A) + 2}$
Indice de Gini	$P(A)(P(B A)^2 + P(\neg B A)^2) + P(\neg A)(P(B \neg A)^2 + P(\neg B \neg A)^2) - P(B)^2 - P(\neg B)^2$
Goodman et Kruskal	$\frac{\sum_i \max_j (P(A_i B_j)) + \sum_j \max_i (P(A_i B_j)) - \max_i (P(A_i)) - \max_j (P(B_j))}{2 - \max_i (P(A_i)) - \max_j (P(B_j))}$
Information mutuelle normalisée	$\frac{\sum_i \sum_j P(A_i B_j) \log_2 \left(\frac{A_i B_j}{P(A_i)P(B_j)} \right)}{-\sum_i P(A_i) \log_2 (P(A_i))}$
J-mesure	$P(AB) \log \left(\frac{P(B A)}{P(B)} \right) + P(A\neg B) \log \left(\frac{P(\neg B A)}{P(\neg B)} \right)$
Support à sens unique	$P(B A) \log_2 \left(\frac{P(AB)}{P(A)P(B)} \right)$
Support à double sens	$P(AB) \log_2 \left(\frac{P(AB)}{P(A)P(B)} \right)$
Variation du support à double sens	$P(AB) \log_2 \left(\frac{P(AB)}{P(A)P(B)} \right) + P(A\neg B) \log_2 \left(\frac{P(A\neg B)}{P(A)P(\neg B)} \right) + P(\neg AB) \log_2 \left(\frac{P(\neg AB)}{P(\neg A)P(B)} \right) + P(\neg A\neg B) \log_2 \left(\frac{P(\neg A\neg B)}{P(\neg A)P(\neg B)} \right)$
Coefficient de corrélation linéaire	$\frac{P(AB) - P(A)P(B)}{\sqrt{P(A)P(B)P(\neg A)P(\neg B)}}$
Platetsky-Shapiro	$P(AB) - P(A)P(B)$
Cosinus	$\frac{P(AB)}{\sqrt{P(A)P(B)}}$
Loevinger	$1 - \left(\frac{P(A)P(\neg B)}{P(A\neg B)} \right)$
Gain d'Information	$\log \frac{P(AB)}{P(A)P(B)}$
Sebag-Schoenauer	$\frac{P(AB)}{P(A\neg B)}$
Moindre contradiction	$\frac{P(AB) - P(A\neg B)}{P(B)}$

4.3. Mesure d'intérêt des règles d'association

Tableau adapté de (Geng et Hamilton, 2006)

4.1.4. Le support d'une règle d'association

Soit deux itemsets X et Y tel que $X \neq \emptyset$, $Y \neq \emptyset$ et $X \cap Y \neq \emptyset$ alors le support de la règle d'association $X \rightarrow Y$, noté $S(X \rightarrow Y)$, représente le nombre de transactions qui contiennent à la fois X et Y . Ainsi, $S(X \rightarrow Y) \Leftrightarrow S(X \cup Y)$. Le calcul du support est donné par l'équation 4.3 où n égal le nombre total de transactions.

$$S(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{n} \quad (4.3)$$

Le support d'une règle d'association est égal au support de l'itemset formé par l'union des items de son antécédent et de son conséquent. À titre d'exemple, $S(\{Am\} \rightarrow \{C, G\}) = S(AMCG) = 0.67$. Le tableau 4.4 donne la mesure de support de quelques règles d'association générées à partir des transactions du tableau 4.1.

Règles	Support
$\{Am\} \rightarrow \{F\}$	$5/9 = 0.56$
$\{F\} \rightarrow \{CB\}$	$2/9 = 0.22$
$\{Am\} \rightarrow \{F, C\}$	$5/9 = 0.56$
$\{Am, F\} \rightarrow \{C\}$	$5/9 = 0.56$
$\{Am\} \rightarrow \{C, G\}$	$6/9 = 0.67$

4.4. Mesure de support

Support calculé pour quelques règles d'association générées à partir des transactions du tableau 4.1.

Le support d'une règle d'association est une mesure importante, car les règles dont le support est très bas sont susceptibles d'apparaître uniquement par hasard (Tan et al., 2006). Par conséquent, l'utilité de ces règles est limitée, voire même nulle. Le support est donc utilisé pour éliminer les règles susceptibles d'être inintéressantes.

4.1.5. La confiance d'une règle d'association

La confiance d'une règle d'association notée $C(X \rightarrow Y)$ représente le nombre de transactions qui contiennent à la fois l'antécédent et le conséquent parmi les transactions qui contiennent l'antécédent.

$$C(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (4.4)$$

La confiance mesure la fiabilité d'une règle d'association. C'est une mesure de précision. Plus la confiance d'une règle est élevée, plus il est probable que le conséquent soit contenu dans les transactions où l'antécédent est présent. Le tableau 4.5 donne la mesure de confiance de quelques règles d'association générées à partir des transactions du tableau 4.1.

Règles	Confiance
$\{Am\} \rightarrow \{F\}$	$5/9 = 0.56$
$\{F\} \rightarrow \{CB\}$	$2/5 = 0.40$
$\{Am\} \rightarrow \{F, C\}$	$5/9 = 0.56$
$\{Am, F\} \rightarrow \{C\}$	$5/5 = 1.00$
$\{Am\} \rightarrow \{C, G\}$	$6/9 = 0.67$

4.5. Mesure de confiance

Mesure de confiance de quelques règles d'association générées à partir des transactions du tableau 4.1.

Déterminer les seuils optimaux n'est pas une tâche simple. Les seuils doivent être fixés de manière à purger les associations causées par du bruit dans les données sans toutefois éliminer une quantité importante d'associations potentiellement intéressantes. Certains travaux relativement récents portent sur la paramétrisation automatique des seuils (Selvi et Tamilarasi, 2011; Selvi et Tamilarasi, 2009). Toutefois, que ce soit pour le support ou la confiance, il n'existe pas de critère unanime permettant de définir le seuil optimal à appliquer à ces mesures. L'extraction des règles d'association demeure un processus exploratoire. Les seuils sont majoritairement fixés par des utilisateurs selon leur connaissance et raffinés en fonction des résultats obtenus.

4.2. L'extraction des règles d'association

Il existe plusieurs algorithmes d'extraction des règles d'associations (Agrawal et al., 1993; Pasquier, 2000; Han et al., 2004). Ces algorithmes comportent sommairement deux étapes. La première est consacrée à la recherche des itemsets fréquents tandis que la seconde est vouée à la production des règles

d'association. Sans stratégie efficace, les calculs nécessaires à l'exécution de ces tâches sont prohibitifs étant donné la quantité de données qui doit être manipulée. À titre de rappel, pour un ensemble d'items de taille d le nombre d'itemsets potentiellement fréquent est 2^d . Ainsi pour seulement 20 items cela représente 1 048 576 itemsets potentiellement fréquents. Le nombre de règles d'association possibles est encore plus considérable. À partir de 20 items, il est possible de générer 3 484 687 250 règles ! Le nombre de règles d'association qu'il est possible de générer à partir de d items peut être calculé à l'aide de l'équation 4.5 (Tan et al., 2006). Considérant que plusieurs milliers d'items peuvent être répartis au sein de transactions, limiter le nombre de combinaisons considérées est une nécessité. L'algorithme APRIORI proposé par Agrawal est l'un des algorithmes les plus utilisés pour extraire les règles d'association (Agrawal et al., 1993). Les deux grandes étapes du processus d'extraction des règles d'association sont exposées dans les sous-sections suivantes en relation avec l'algorithme APRIORI.

$$R = 3^d - 2^{d+1} + 1 \quad (4.5)$$

4.2.1. La recherche des itemsets fréquents

La première étape du processus d'extraction des règles d'association consiste à rechercher les itemsets fréquents. Le nombre de règles d'association non trivial qu'il est possible de générer peut être contrôlé en fonction du nombre d'itemsets fréquents. En effet, plus le nombre d'itemsets fréquents est petit, plus le nombre de règles d'association est limité. Le problème de l'effort computationnel

nécessaire à l'extraction des règles d'association se réduit donc au problème de l'optimisation de la recherche des itemsets fréquents (Pasquier, 2000). Afin d'optimiser la recherche des itemsets fréquents, l'algorithme APRIORI propose un mécanisme itératif. Les différentes itérations visent à repérer les itemsets fréquents. Les itemsets sont préalablement ordonnés de manière à faciliter cette opération. Pour ce faire, que ce soit pour l'algorithme APRIORI ou pour l'une de ces nombreuses variantes, l'ordre lexicographique des caractères est généralement appliqué (Pasquier, 2000). Cet ordre permet d'éviter des calculs redondants. Le treillis des itemsets est parcouru un niveau à la fois, des 1-itemsets aux itemsets de taille maximale. À la première itération, tous les itemsets candidats de taille 1 sont générés et comparés aux itemsets contenus dans les diverses transactions de manière à pouvoir calculer leur support. Les 1-itemsets candidats inféquents sont rejetés tandis que ceux fréquents sont conservés. À la deuxième itération, les itemsets candidats de taille 2 sont générés à partir des 1-itemsets fréquents découverts lors de l'itération précédente. Ces nouveaux candidats sont leurs tours comparés aux itemsets contenus dans les diverses transactions. Des itérations sont effectuées jusqu'à ce qu'aucun nouvel itemset fréquent ne soit découvert. Ce procédé permet de réduire considérablement le nombre d'itemsets à traiter. Le nombre d'itérations maximum est $k_{max} + 1$ où k_{max} correspond à la taille maximale des itemsets fréquents. L'exemple 4.2 donne une portion du pseudo-code de l'algorithme APRIORI.

```

k = 1;
Fk = I;
tant que (Fk-1 ≠ 0) faire
    k = k + 1;
    Ck = aprioriGen(Fk);
    pour tous les transaction t ∈ T faire
        Ct = subset(Ck, t);
        pour tous les candidats c ∈ Ct
            σ(c) ++;
        fin pour
    fin pour
    Fk = {c ∈ Ck | σ(c) ≥ minsup};
fin tant que
itemsets fréquents =  $\bigcup_k F_k$ 

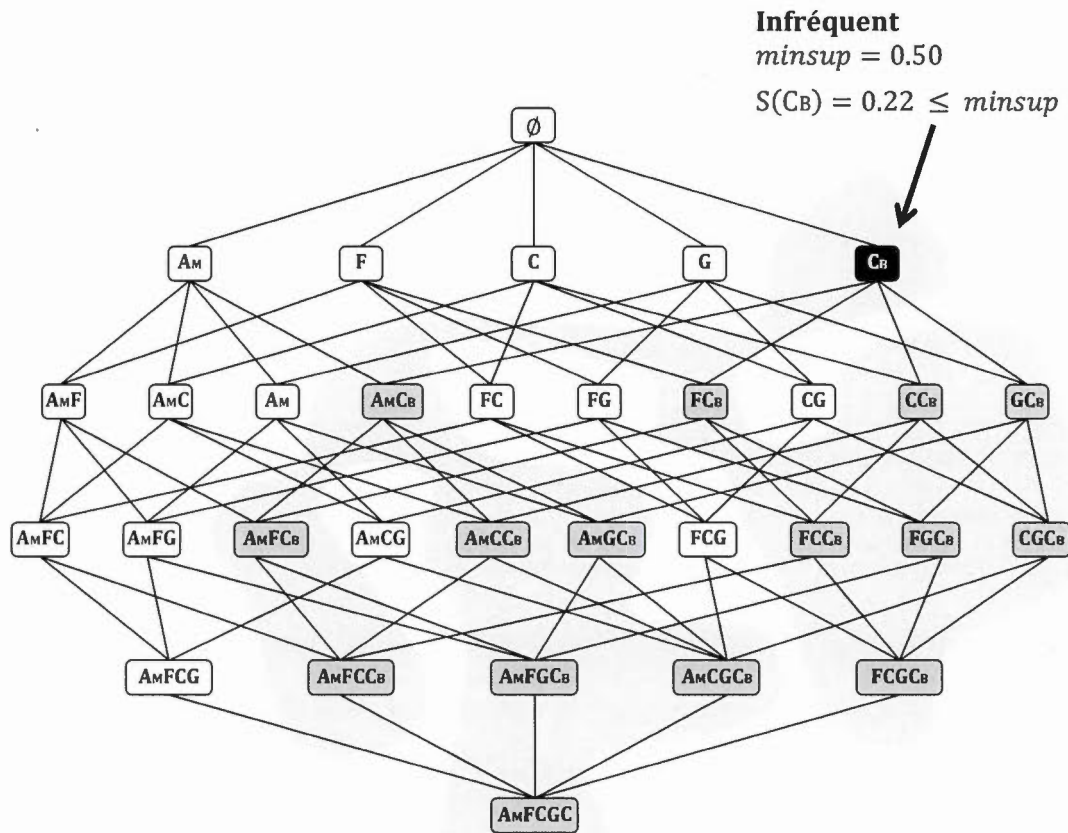
```

Exemple 4.2 : Pseudo-code de l'algorithme APRIORI qui permet d'obtenir les itemsets fréquents. La fonction **aprioriGen** est responsable de retourner tous les *k*-itemsets candidats qu'il est possible de générer à partir de l'ensemble passé en paramètre tandis que la fonction **subset** est responsable de retourner les itemsets candidats qui sont contenus dans une transaction donnée.

Le principe sur lequel repose l'algorithme APRIORI est que le support d'un itemset n'est jamais supérieur au support des itemsets qu'il renferme. Cette propriété est nommée anti-monotonie (équation 4.6).

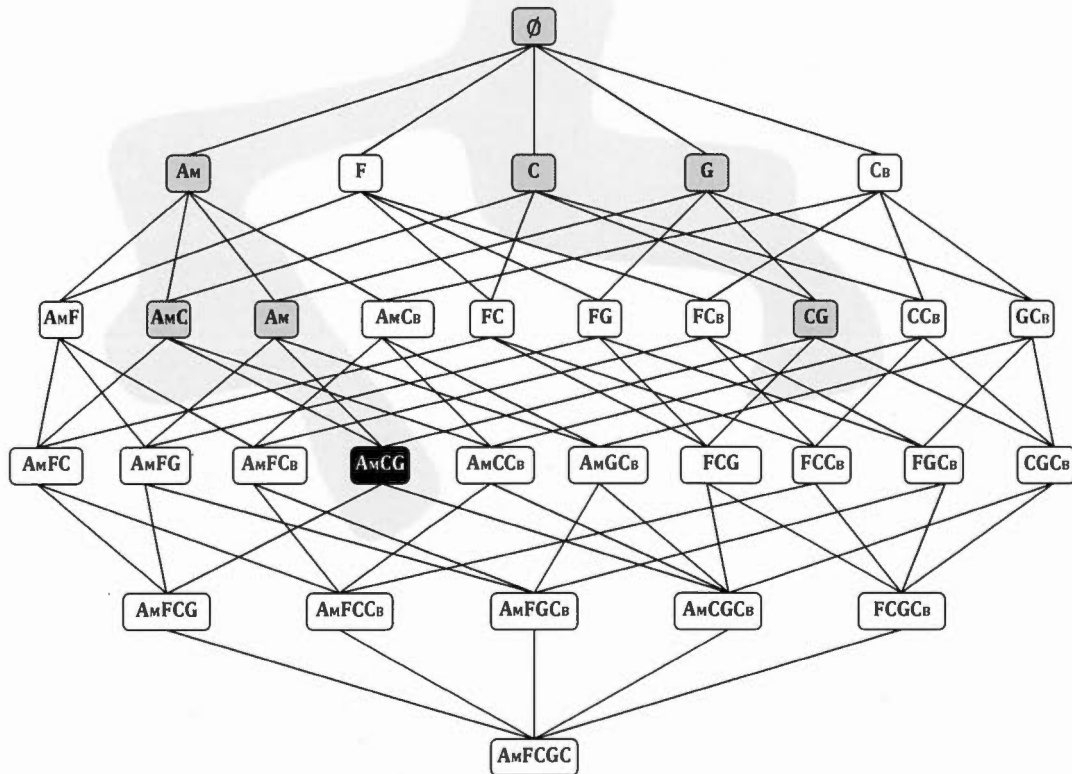
$$\textbf{Propriété 1 :} \quad \forall X, \forall Y : (X \subseteq Y) \Rightarrow S(X) \geq S(Y) \quad (4.6)$$

Afin de démontrer comment l'anti-monotonie peut servir à élaguer le treillis d'itemsets, considérons les itemsets du tableau 4.2 et le treillis illustré à la figure 4.1. L'itemset {CB} possède 15 descendants : { AM, CB }, { F, CB }, { C, CB }, { G, CB }, { AM, F, CB }, { AM, C, CB }, { AM, G, CB }, { F, C, CB }, { F, G, CB }, { C, G, CB }, { AM, F, C, CB }, { AM, F, G, CB }, { AM, C, G, CB }, { F, C, G, CB } et { AM, F, C, G, CB }. Considérant que { CB } est infrequent alors ses 15 descendants peuvent être élagués puisque $\forall X, Y : (X \subseteq Y) \Rightarrow S(X) \geq S(Y)$. Au total, 16 combinaisons peuvent être éliminées ce qui représente 50% des combinaisons possibles. La figure 4.2 illustre l'impact de ce nettoyage sur le treillis.



4.2. Élagage de l'itemset non fréquent C_B et de ses descendants

L'anti-monotonie peut également servir à déduire que des itemsets sont fréquents. Considérons trois itemsets non vides X, Y, Z tel que $Z \subset Y \subset X$, si X est fréquent alors Y et Z le sont également, car $S(X) \leq S(Y) \leq S(Z)$. Par exemple, sachant que $\{A_M, C, G\}$ est fréquent alors tous ses sous-ensembles soit $\{A_M\}$, $\{C\}$, $\{G\}$, $\{A_M, C\}$, $\{A_M, G\}$ et $\{C, G\}$ sont alors également fréquents. L'itemset $\{\emptyset\}$ n'étant pas utile il est ignoré. La figure 4.3 illustre cette relation.



4.3. Itemsets fréquents déduits de $\{Am, C, G\}$

Dans la majorité, voir la totalité des cas, l'itemset $\{\emptyset\}$ n'est pas utile, il est donc ignoré.

4.2.2. La production des règles d'association

La deuxième étape du processus d'extraction des règles d'association consiste à produire les règles. Cette étape ne nécessite pas de balayage des transactions, car les règles sont générées à partir des itemsets fréquents découverts lors de la première étape du processus. Le support d'une règle d'association est identique au support de l'itemset fréquent à partir duquel elle est générée (Tan et al., 2006). Ainsi toutes les règles produites à partir d'itemsets fréquents ont un support supérieur ou égal au support minimum établi. Il est possible de générer jusqu'à $2^k - 2$ règles d'association à partir d'un k -itemset fréquent. Cette relation entre la taille d'un itemset et le nombre de règles d'association qu'il est possible de générer à partir de cet itemset est donnée par l'équation 4.7. Tout comme lors de la recherche des itemsets fréquents, l'application d'une stratégie est requise afin de réduire le nombre de combinaisons possibles en éliminant les combinaisons présumées inintéressantes.

$$R' = 2^k - 2 \quad (4.7)$$

Considérons l'itemset fréquent $Y = \{A_m, C, G\}$ et un de ses sous-ensembles $X = \{A_m\}$, alors le sous-ensemble $(Y - X) = \{C, G\}$. Selon la définition de la confiance tel que donnée à l'équation 4.4, la confiance de la règle $X \rightarrow (Y - X)$ est égale à $\sigma(X \cup (Y - X)) / \sigma(X)$. Ainsi, $C(X \rightarrow Y - X) = \sigma(A_m, C, G) / \sigma(A_m)$ et par conséquent $C(X \rightarrow Y - X)$ est égal au rapport $S(Y) / S(X)$. L'algorithme APRIORI utilise la mesure de confiance pour limiter le nombre de règles produites. Pour ce faire, pour chaque itemset fréquent Y , tous les sous-ensembles X tels que $X \subset Y$ sont d'abord générés. Le rapport $S(Y) / S(X)$ est

ensuite calculé. Si ce rapport est supérieur ou égal à un seuil minimum de confiance alors la règle d'association $X \rightarrow Y - X$ est créée. Tout comme pour le support, la confiance possède une propriété qui permet d'éliminer certaines combinaisons. En effet, si la confiance d'une règle $X \rightarrow Y - X$ est en deçà du seuil minimum de confiance, alors la confiance de toutes les règles $X' \rightarrow Y - X'$ où $X' \subset X$ est également en dessous de ce seuil. Cette propriété est donnée par l'équation 4.8.

Propriété 2 :
$$C(X' \rightarrow Y - X') \leq C(X \rightarrow Y - X) \text{ où } X' \subset X \quad (4.8)$$

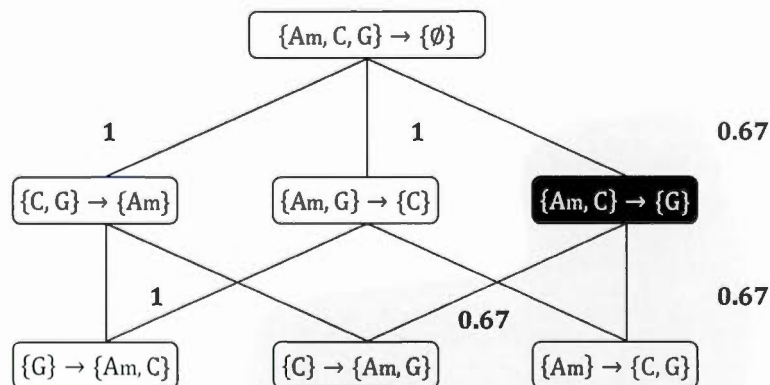
Considérons deux règles d'association, $X \rightarrow Y - X$ et $X' \rightarrow Y - X'$ où $X' \subset X$. La confiance de la règle $X' \rightarrow Y - X'$ est égale au rapport $S(Y) / S(X')$ tandis que la confiance de la règle $X \rightarrow Y - X$ est égale au rapport $S(Y) / S(X)$. Puisque $X' \subset X$ alors $S(X') \geq S(X)$ et par conséquent $C(X' \rightarrow Y - X') \leq C(X \rightarrow Y - X)$. Ainsi, lorsque la confiance d'une règle d'association est sous le seuil minimum de confiance alors toutes les autres règles qui contiennent le conséquent de cette règle peuvent être éliminées. La figure 4.4 illustre un exemple possible de réduction basée sur les données des tableaux 4.6 et 4.7.

Candidats	Support
{Am, C, G}	6
{Am, C}	9
{Am, G}	6
{C, G}	6
{A}	9
{C}	9
{G}	6

4.6. Support des sous-ensembles candidats

Règles candidates	Confiance
$\{C, G\} \rightarrow \{Am\}$	$6/6 = 1.00$
$\{Am, G\} \rightarrow \{C\}$	$6/6 = 1.00$
$\{Am, C\} \rightarrow \{G\}$	$6/9 = 0.67$
$\{G\} \rightarrow \{Am, C\}$	$6/6 = 1.00$
$\{C\} \rightarrow \{Am, G\}$	$6/9 = 0.67$
$\{Am\} \rightarrow \{C, G\}$	$6/9 = 0.67$

4.7. Confiance des règles candidates



4.4. Nettoyage des règles d'association

Puisque $C(\{Am, C\} \rightarrow \{G\})$ est inférieure au seuil de confiance minimum alors les règles $\{Am, C\} \rightarrow \{G\}$ et $\{Am, C\} \rightarrow \{G\}$ peuvent être éliminées.

Différentes méthodes ont été proposées pour optimiser le processus d'extraction des règles d'association. Comme la recherche des itemsets fréquents est la phase la plus coûteuse du processus, ces alternatives portent sur l'amélioration de cet aspect (Tan et al., 2004 ; Pasquier, 2000)

4.3. Les règles d'association et la musique

L'utilisation des règles d'association dans le domaine du MIR est relativement récente. Swathi et Reddy ont proposé d'utiliser les règles d'associations afin d'améliorer les systèmes de recommandation de musique (Swathi et Reddy, 2014). Plusieurs systèmes de recommandation utilisent le filtrage collaboratif comme stratégie. Dans ce contexte, les préférences musicales des utilisateurs sont utilisées pour leur suggérer des pièces musicales susceptibles de les intéresser. La performance de ces systèmes dépend grandement de la connaissance préalable des préférences des utilisateurs. L'approche proposée par Swathi et Reddy consiste à utiliser les règles d'association lorsque cette connaissance est limitée. C'est en combinant un nombre restreint de préférences et des règles d'associations fréquentes que le système développé par Swathi et Reddy parvient à recommander de nouvelles pièces. Ainsi, une pièce est recommandée lorsqu'elle est contenue dans le conséquent d'une règle et que les pièces qui forment l'antécédent de cette règle figurent parmi les préférences de l'utilisateur. L'approche proposée est intéressante. Cependant, les auteurs donnent peu de détail sur les performances obtenues.

Conklin est l'un des premiers à avoir proposé d'utiliser les règles d'association pour classifier des données musicales (Conklin, 2009). La solution suggérée par Conklin consiste à utiliser les règles d'association d'une part pour extraire des motifs séquentiels inscrits dans les mélodies et d'autre part pour établir des

relations possibles entre ces motifs et différents genres musicaux. L'identification de la mélodie qui compose une musique polyphonique est une tâche complexe. Pour cette raison, l'approche proposée par Conklin se limite au traitement de fichiers symboliques (p. ex. le format de fichier MIDI). La particularité des fichiers symboliques est qu'ils ne sont pas composés des points d'échantillonnage qui traduisent la forme de l'onde sonore. Les formats symboliques contiennent une description détaillée de la musique. Ils fournissent explicitement les informations concernant le ton, l'intensité et la durée des notes qui composent une mélodie. Les résultats des expérimentations effectuées par Conklin suggèrent que les règles d'association peuvent servir à dégager des motifs dans les mélodies qui sont suffisamment discriminants pour soutenir un processus de classification. Selon Conklin, une classification cohérente à 77% peut être obtenue à partir des motifs dégagés par les règles d'association. Toutefois, ces travaux relèvent également une limitation. Dans certains cas, les règles d'association ne permettent pas de dégager des motifs discriminants. Par conséquent, certains fichiers ne peuvent être classifiés.

Arjannikov et Zhang se sont également intéressés à l'utilisation des règles d'association pour classifier des données musicales. Cependant, plutôt que d'opter pour l'usage de documents symboliques, leurs travaux ont porté sur la classification de documents polyphoniques. Ils ont proposé une méthode basée sur l'utilisation des règles d'association pour prédire le genre d'un ensemble de pièces musicales (Arjannikov et Zhang, 2014). Leur modèle, à l'instar de plusieurs autres voués à cette tâche, utilise des caractéristiques de l'onde sonore comme descripteurs. Des règles d'association sont générées à partir de ces caractéristiques et utilisées pour classifier les données. L'approche préconisée est supervisée. Leur classifieur est entraîné afin de reconnaître les règles les plus représentatives des genres étudiés.

L'usage des règles d'association est d'un grand intérêt, car ces règles peuvent être adaptées pour faire ressortir des connaissances cachées dans les descripteurs qui traduisent la musique. Contrairement aux réseaux de neurones qui possèdent également cette capacité, elles ne sont pas opaques (Tada et al., 2010). Par conséquent, il est plus facile de les interpréter et au besoin de les adapter. Ainsi, les règles d'associations peuvent facilement manipuler des descripteurs utilisés pour exprimer différentes perspectives. Cette approche demeure aujourd'hui peu explorée.

CHAPITRE V

SOLUTION PROPOSÉE AU PROBLÈME D'EXPLORATION

Comme il a été présenté dans les chapitres précédents et de manière plus approfondie dans les chapitres 2 et 3, la musique peut être abordée selon différentes perspectives. Chacune des perspectives participe à définir comment la musique est perçue. C'est à travers le traitement de ces perspectives que se crée l'expérience musicale. À la lumière de ce constat, il paraît évident que l'exploration de données musicales doit être abordée selon différentes perspectives.

Sur le plan informatique, le développement d'un système d'aide à l'exploration de données musicales pose principalement deux grandes difficultés. Les différentes perspectives à partir desquelles la musique peut être abordée sont représentées par des descripteurs appartenant à différents niveaux d'abstraction. Il existe une multitude de techniques pour extraire et manipuler ces descripteurs. Les descripteurs de bas niveau sont créés à partir des données brutes contenues dans les documents audio. Le taux de passage par zéro, les coefficients spectraux et l'étalement fréquentiel peuvent servir à créer ces descripteurs. Les descripteurs utilisés pour décrire la musique peuvent également appartenir à des niveaux moins abstraits. Par exemple, les descripteurs peuvent être définis à partir des données textuelles périphériques aux documents audio ou encore à partir d'informations factuelles concernant le profil des utilisateurs, leurs préférences et leurs motivations. La difficulté liée à

l'extraction des descripteurs n'est pas située au niveau des mécanismes d'extraction mais plutôt dans la sélection des descripteurs à considérer. L'usage d'un descripteur peut être inefficace dans un contexte et d'une grande utilité dans un autre.

La deuxième difficulté est que l'exploration des données musicales requiert la mise en œuvre de diverses stratégies et le développement de chacune de ces stratégies est un sujet de recherche en soi. La classification seule ne suffit pas à couvrir l'ensemble de ce qu'est l'exploration de données musicales. En effet, l'exploration de données musicales peut se faire sur un ensemble de documents sonores mais également à l'intérieur des documents eux-mêmes. Ainsi, l'exploration de données musicales peut devoir requérir des techniques associées à l'analyse de la structure des pièces musicales par exemple. Un autre aspect qui doit être considéré lors de la conception d'un système d'aide à l'exploration de données musicales est l'accès aux documents. Des mécanismes doivent être offerts pour encourager la prospection des documents. C'est particulièrement cet aspect que couvrent les techniques de recommandations.

5.1. Solution proposée

Il est proposé d'utiliser les règles d'association pour répondre aux deux difficultés liées au développement d'un système d'aide à l'exploration de données musicales. L'absence ou la présence de certains descripteurs est un indice permettant de cibler la nature d'une musique. Il est soutenu que les relations entre ces mêmes descripteurs constituent des indices encore plus riches. Les règles d'association expriment des relations entre des items qui cooccurrent fréquemment. En plus de mesurer les fréquences d'apparition, elles

rehaussent les liens entre les items. Elles indiquent spécifiquement les combinaisons qui définissent un ensemble de données. L'examen de ces combinaisons peut servir d'une part à identifier les descripteurs les plus discriminants et d'autre part à comprendre comment ces descripteurs sont agencés au sein des données. Une autre propriété souhaitable que possèdent les règles d'associations est qu'elles sont flexibles. Elles peuvent être appliquées à des données dont la nature peut varier. Il est donc possible d'envisager de les appliquer à des descripteurs appartenant aux différents niveaux d'abstraction. Les règles d'association peuvent être adaptées à plusieurs usages. En plus de leur remarquable capacité à dégager des connaissances cachées dans de vastes ensembles de données, leur utilité a été démontrée pour assister l'interprétation de classes de similarité (Biskri et al., 2012), reformuler des requêtes (Biskri et Rompré, 2012), classifier des données (Conklin, 2009; Arjannikov et Zhang, 2014) et pour recommander des musiques (Swathi et Reddy, 2014). Ainsi, la même approche peut servir à la fois à identifier les caractéristiques les plus discriminantes et supporter la mise en œuvre de différentes stratégies pour soutenir l'exploration de données musicales.

L'utilisation des règles d'association est cependant sujette à une problématique importante. Les résultats qu'elles engendrent dépendent du nombre de variables considérées. Des données disparates peuvent entraîner une explosion combinatoire (Guillaume et Khenchaf, 2000). Plusieurs règles peuvent être générées et plusieurs de ces règles peuvent être redondantes ou peu significatives. Cette condition peut provoquer des résultats indésirables particulièrement lorsque les règles d'association sont utilisées à des fins de classification. Des règles non significatives peuvent se retrouver en abondance dans plusieurs documents. Cette présence est susceptible d'engendrer la création de faux liens entre des documents. Une autre complication que

provoque l'extraction massive de règles d'association est qu'elle peut causer une surcharge computationnelle. Ces enjeux peuvent toutefois être résolus en exerçant un contrôle sur le nombre de variables considérées.

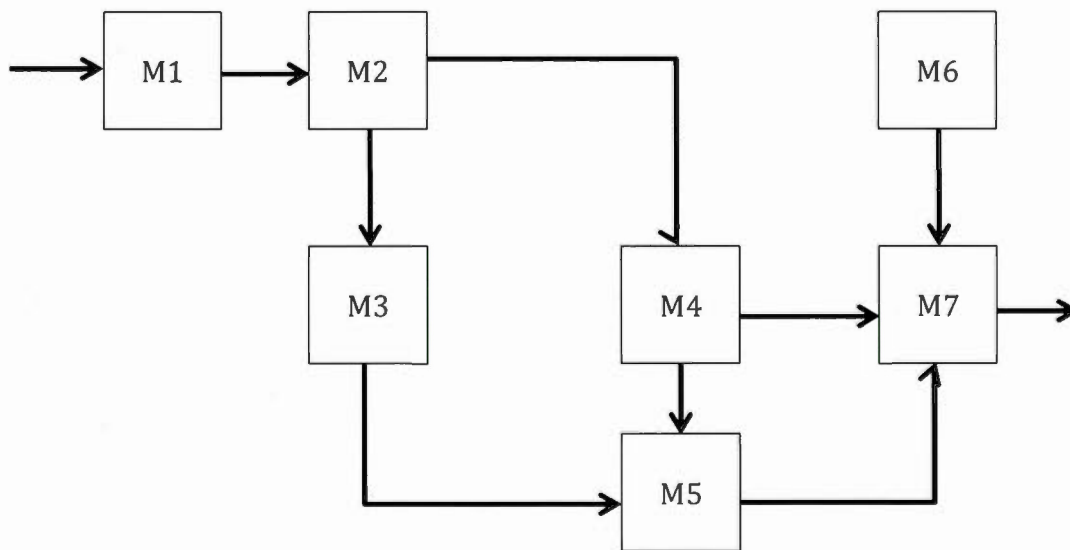
5.2. Considérations architecturales

Une partie importante du travail effectué lors de ce projet de recherche a été consacré au développement de plusieurs chaînes de traitement. Une chaîne de traitement est une succession de plusieurs fonctions. Si les fonctions qui composent une chaîne de traitement sont computationnelles alors cette chaîne de traitement est présumée computationnelle également. Quatre types d'opérations sont impliquées (Biskri et Meunier, 2002; Meunier et al., 2005) :

1. La réception des intrants informationnels
2. Le traitement des intrants (p. ex. transformation, catégorisation, etc.)
3. Le contrôle
4. La diffusion

Les opérations impliquées dans la création d'une chaîne de traitement peuvent être réalisées en parallèle ou en série. Chacun des modules est relayé à l'exécution d'une opération précise. Les multiples opérations sont ainsi encapsulées. Selon Pressman (Pressman, 2001), l'encapsulation procure trois grands avantages. Premièrement, elle réduit le risque de propagation d'effet non désirable lors de modification. Cette propriété est grandement désirable pour suivre l'évolution des techniques liées au traitement du signal. Deuxièmement, l'encapsulation favorise la réutilisation. Même s'il existe plusieurs approches pour traiter les dimensions musicales, plusieurs d'entre elles exploitent des

algorithmes communs. Par exemple, la transformée de Fourier est une opération quasi incontournable pour le traitement du signal sonore (Cavicchi, 1999). Ainsi, un module dédié à cette opération peut être réutilisé dans plusieurs chaînes de traitements qui s'intéressent au contenu fréquentiel de l'onde sonore. Finalement, l'encapsulation facilite l'interfaçage entre les diverses opérations. Une opération qui utilise le résultat d'une autre opération n'est pas contrainte à connaître le détail des structures internes à cette opération. Par conséquent, le couplage entre les différentes opérations tend à être réduit. Un faible couplage réduit les interdépendances entre les modules ce qui accroît leur flexibilité. Face à la multiplicité des interventions requises pour décortiquer, analyser et interpréter la musique, la robustesse, la réutilisation et la flexibilité sont des propriétés recherchées lors du développement d'un système d'exploration de données musicales.



5.1. Architecture modulaire

Dans une architecture modulaire, plusieurs modules collaborent à la création de chaînes de traitements. Chacun des modules est responsable de l'exécution d'une tâche spécifique. Une chaîne de traitement peut être rehaussée en lui ajoutant des modules. Les modules doivent être conçus de manière à ce qu'ils puissent être exploités dans différentes chaînes de traitement.

5.3. Architecture proposée

Selon les hypothèses émises en introduction, une architecture dédiée à l'exploration de données musicales doit au minimum :

- Permettre le traitement de plusieurs types d'information dont le niveau d'abstraction peut varier.

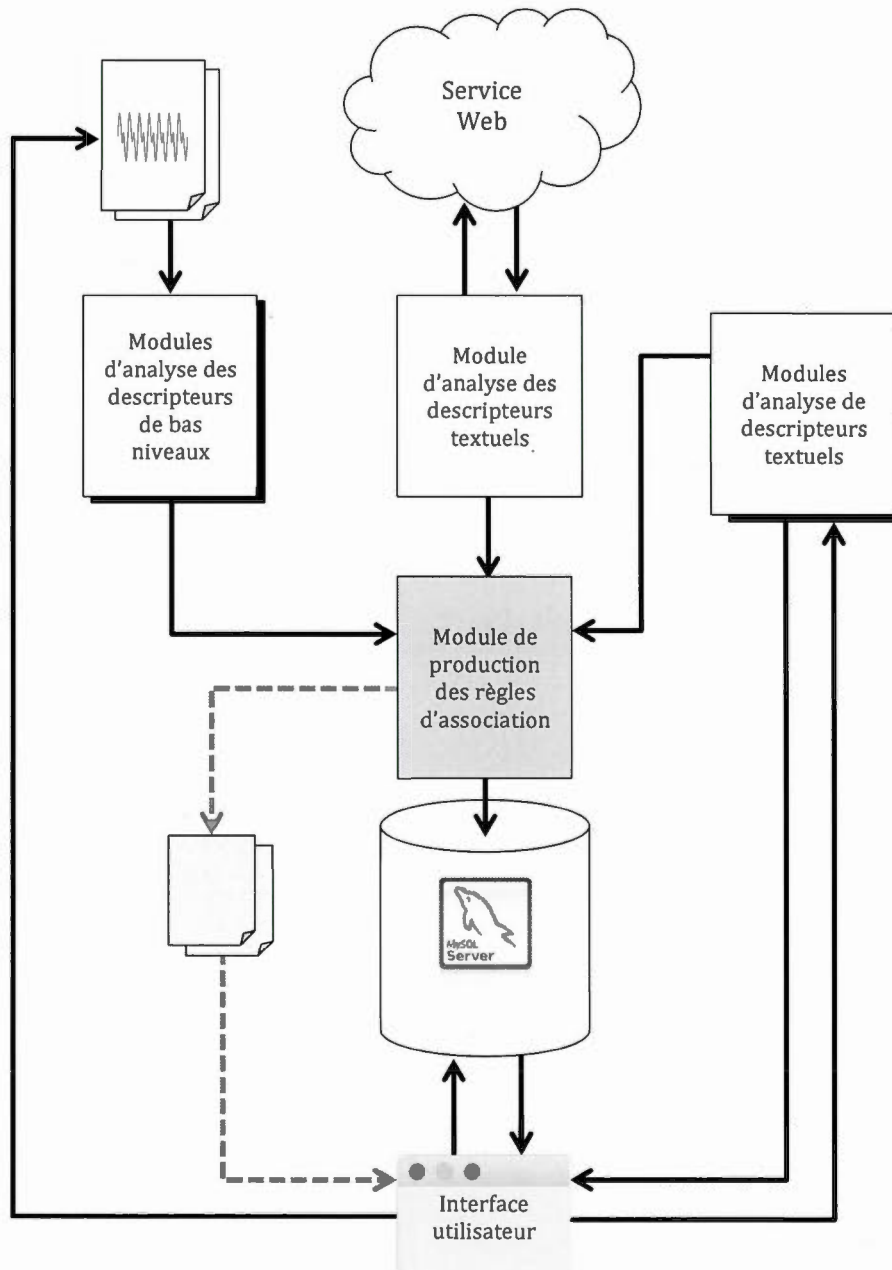
Il est supposé que l'expérience que procure l'écoute d'une musique dépend du traitement d'informations multidimensionnelles. Une architecture dédiée au traitement de données musicales doit donc permettre de répondre à cette particularité.

- Faciliter la détection d'associations fortes entre les descripteurs utilisés pour décrire la musique.

L'une des hypothèses posées est que les relations qui existent entre les descripteurs sont tout aussi importantes que la présence ou l'absence de ces descripteurs.

- Favoriser la mise en œuvre de plusieurs stratégies pour soutenir l'exploration de données musicales.

L'exploration de données musicales peut être améliorée lorsque des techniques associées à la classification, à l'analyse de la structure des pièces musicales et à la recommandation sont utilisées conjointement.

**5.2. Architecture proposée**

La figure 5.2. illustre l'architecture proposée pour répondre aux trois contraintes citées précédemment. Elle repose sur l'intégration d'un module dédié à l'extraction des règles d'association. Afin d'éviter de devoir développer plusieurs variantes de ce module, il est proposé de convertir tous les intrants en chaînes de caractères. Même si ce choix conceptuel ne permet pas d'optimiser la gestion de la mémoire, il offre la flexibilité recherchée. En effet, les chaînes de caractères sont en mesure de contenir à la fois des nombres et du texte. L'extraction des règles d'association est une opération qui peut demander un temps considérable. Pour cette raison, il est suggéré que cette tâche soit effectuée en arrière-plan et que les règles extraites soient stockées dans une base de données ou dans des fichiers de sauvegarde afin de demeurer accessibles. La modularité est un aspect important de l'architecture préconisée. Il est proposé que les transactions soumises en input au module d'extraction des règles d'association soient générées à partir de modules spécialisés dans le traitement de descripteurs spécifiques. Un module doté de fonctionnalités liées au traitement de l'onde sonore peut être utilisé pour générer les transactions composées de descripteurs de bas niveau par exemple. Les données périphériques aux documents audio ont tendance à évoluer dans le temps. C'est le cas notamment pour la discographie des artistes. Afin d'assurer une mise à jour de ces informations, il est proposé de collecter les données périphériques aux enregistrements sonores à partir d'appels à des services web (p. ex. Echonest¹⁸ et Amazon API¹⁹). C'est à partir d'interaction avec l'utilisateur que ses préférences et ses motivations peuvent être capturées. L'architecture proposée préconise l'ajout de nouveaux modules afin de permettre l'inclusion de nouveaux descripteurs.

¹⁸ <http://the.echonest.com/>

¹⁹ <http://aws.amazon.com/fr/documentation/>

5.4. Spécification du prototype développé

Le travail effectué dans le cadre de cette thèse a mené au développement d'un prototype. Comme l'objectif général de cette thèse est de démontrer la flexibilité et la portabilité des règles d'association dans un contexte d'aide à l'exploration multidimensionnelle de données musicales, seules quelques portions de l'architecture ciblée ont été développées. L'accent a été mis sur le développement des composantes permettant de démontrer les capacités des règles d'association.

Le prototype a été développé sous la plateforme Mac OS X à l'aide de l'éditeur XCode version 6.2. Le code source du prototype a été écrit en langage objective-c. Ces choix ont été motivés par l'accès aux bibliothèques *AudioToolbox Framework*²⁰ et *Accelerate Framework*²¹ dédiées respectivement à la manipulation des données audio et au traitement du signal (Admanson et Avila, 2012). Ces bibliothèques sont offertes gratuitement par Apple. La bibliothèque *AudioToolbox Framework* supporte un vaste choix de formats de fichiers audio, dont le format MP3. La bibliothèque *Accelerate Framework* offre, quant à elle, des méthodes facilitant l'application de la transformée de Fourier.

5.4.1. Stockage des données

Les règles d'association et les informations complémentaires retournées par les services web peuvent être stockées dans une base de données. La base de

²⁰<https://developer.apple.com/library/ios/documentation/MusicAudio/Reference/CAAudioToolboxRef/>

²¹https://developer.apple.com/library/ios/documentation/Accelerate/Reference/AccelerateFWRef/index.html#//apple_ref/doc/uid/TP40009465

données retenue lors du développement du prototype est MySQL²². Le choix de cette base de données a été motivé par le fait qu'elle peut être utilisée sous une licence GPL²³ c'est-à-dire qu'elle est libre de droits. Les données peuvent également être conservées sur des fichiers de sauvegarde afin d'éviter d'avoir recours à la base de données.

5.4.2. Formats de fichier audio supportés

La librairie *AudioToolbox Framework* utilisée dans le développement du prototype permet la lecture de plusieurs formats de fichier comme les formats MP3, WAVE et AIFF. Le format MP3 est le format de fichier audio qui est le plus utilisé. La majorité des données musicales qui transigent sur les réseaux informatiques sont représentées sous ce format. Pour cette raison, un système d'aide à l'exploration de données musicales se doit de supporter ce format de fichier. Le format MP3 se distingue des autres par le niveau de compression qu'il offre. Les fichiers audio non compressés sont volumineux. Il faut jusqu'à dix méga-octets pour contenir 1 minute de musique non compressée. À titre de comparaison, un fichier MP3 de cette taille est en mesure de contenir une pièce musicale d'environ 4 minutes. Sommairement, le taux de compression est obtenu en supprimant certaines fréquences jugées inaudibles. Les fréquences écartées sont d'une part les fréquences situées en dehors de l'intervalle des fréquences que l'oreille humaine est en mesure de percevoir et d'autre part les fréquences jugées masquées. Une fréquence est dite masquée lorsque son niveau sonore est très faible par rapport au niveau sonore d'une autre fréquence contenue dans le signal. La principale critique envers le format de fichier MP3 provient de cette réduction du spectre fréquentielle. Malgré cette critique,

²²<http://dev.mysql.com/downloads/mysql/>

²³<http://www.gnu.org/licenses/licenses.html>

l'altération du spectre fréquentiel créée par la compression des données est à peine perceptible pour plusieurs d'entre nous. De plus, le format MP3 est en mesure de livrer l'essentiel du message sonore. Une musique peut être écoutée, reconnue et appréciée.

Peu importe le format de fichier soumis en input, les données sont toujours converties à l'aide de la librairie *AudioToolbox Framework* dans le format linéaire PCM mono afin de simplifier les traitements.

CHAPITRE VI

EXPÉRIMENTATIONS ET DISCUSSION

Plusieurs expérimentations ont été menées afin de vérifier comment les règles d'association peuvent être employées pour effectuer une exploration multiniveau de données musicales. La revue de la littérature présentée dans les chapitres 2 et 3 révèle que le genre musical est positionné au centre de plusieurs stratégies vouées à faciliter l'exploration de données musicales. Le genre est un descripteur qui appartient au niveau d'abstraction intermédiaire. La première expérimentation a été effectuée afin de pouvoir mesurer les limites lorsque les descripteurs considérés traduisent une seule perspective. Lors de cette expérimentation, un examen de trois schémas de classification basés sur le genre musical a été fait. Des similarités et des contrastes ont été observés dans les modèles étudiés. Dans la mesure où il existe des limites à considérer des descripteurs appartenant à un seul niveau d'abstraction, il est d'un grand intérêt de chercher un moyen pour repousser ces limites. Il est présumé que c'est entre autres par l'ajout de mécanismes capables de manipuler des descripteurs appartenant à différents niveaux d'abstraction que ces limites peuvent être repoussées. Les expérimentations II, III et IV ont été réalisées afin de vérifier la capacité des règles d'association d'une part à manipuler des descripteurs appartenant à différents niveaux d'abstraction et d'autre part à faciliter la mise en œuvre de divers mécanismes d'aide à l'exploration de données musicales. La deuxième expérimentation réalisée a porté sur l'évaluation de la capacité des règles d'association à manipuler des descripteurs de bas niveaux. Afin de

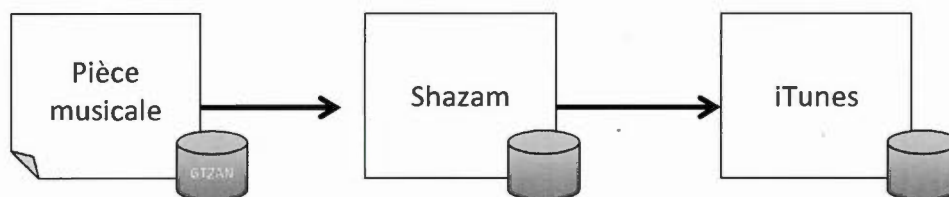
permettre cette évaluation, un modèle de classification a été développé et testé. La troisième expérimentation a été consacrée à vérifier la portabilité des règles d'association. L'approche exploitée lors de l'expérimentation II a été adaptée afin de permettre l'analyse de pièces musicales. Finalement, la quatrième expérimentation a été effectuée pour démontrer que les règles d'association pouvaient être appliquées à des descripteurs autres que des descripteurs de bas niveaux. Les règles d'association ont été utilisées pour bonifier la recherche d'artistes similaires à un artiste cible. Les résultats des expérimentations effectuées sont présentés dans les sections suivantes.

6.1. Expérimentation I : évaluation de classifications existantes

La plupart du temps, les mécanismes d'aide à l'exploration de données musicales s'articulent autour de l'usage de descripteurs appartenant au niveau intermédiaire. Parmi ces descripteurs, le genre musical est le plus commun. Dans le cadre de cette thèse, il est présumé que le fait de considérer des descripteurs appartenant à un seul niveau d'abstraction limite l'exploration de données musicales. Dans l'objectif de mesurer ces limites, l'évaluation de 3 différents schémas de classification a été effectuée. Les modèles considérés ont été sélectionnés de manière à pouvoir représenter différents points de vue. Le premier modèle, celui de Tzanetakis, représente un schéma de classification fréquemment employé dans la recherche dédiée au MIR. Le second modèle, Shazam, représente un schéma hybride cité à la fois dans la littérature et adapté à l'industrie (Wang, 2003). Finalement, le troisième modèle, iTunes, représente le schéma utilisé par le plus grand détaillant de musique en ligne.

6.1.1. Méthodologie

Cette expérimentation a également comme objectif de mesurer les similarités et les contrastes entre les 3 modèles étudiés afin de définir un cadre de référence. Pour ce faire, un ensemble témoin a été créé en sélectionnant des pièces dans l'ensemble GTZAN²⁴. L'une des particularités de cet ensemble de données est que seul le genre présumé est connu. Une fois l'ensemble témoin formé, le titre et le nom de l'interprète de chacune des pièces ont été récupérés à l'aide de l'application Shazam. Le choix de cette application n'est pas dicté par un souci de performance mais plutôt parce qu'elle est l'une des applications les plus employées en reconnaissance de musique. En effet, le nombre d'utilisateurs mensuels de Shazam est estimé à 100 millions²⁵. Le genre retourné par Shazam a été collecté pour former le deuxième modèle. Finalement, le titre de la pièce et le nom de l'interprète ont été utilisés pour effectuer une recherche sur iTunes. Le genre donné par iTunes a été collecté à son tour et utilisé pour construire le troisième modèle. Cette séquence d'opérations constitue un exemple typique d'exploration de données musicale à partir d'appareils mobiles.



6.1. Séquence des opérations de l'expérimentation I

²⁴http://marsyasweb.appspot.com/download/data_sets/

²⁵<http://news.shazam.com/>

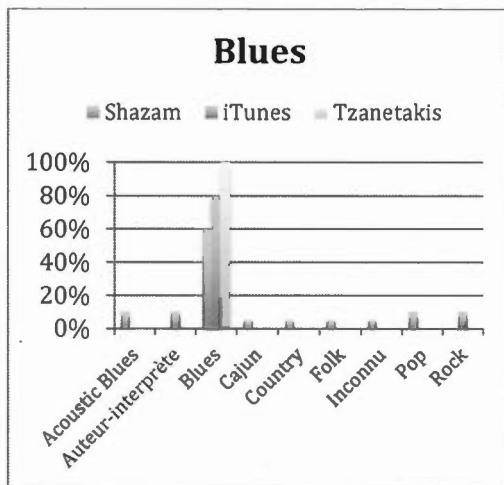
6.1.2. Ensemble de données

L'ensemble témoin a été créé en sélectionnant aléatoirement des pièces dans la collection GTZAN. Cette collection de données est composée de documents sonores rassemblés par George Tzanetakis et utilisés initialement dans le cadre de l'article « *Musical Genre Classification of Audio Signals* » (Tzanetakis et Cook, 2002). Dix genres musicaux sont représentés : le blues, le classique, le country, le disco, le hip-hop, le jazz, le métal, la musique populaire, le reggae et le rock. Chacun des genres est représenté par 100 extraits de 30 secondes. Les extraits sont identifiés par le nom du genre qu'ils représentent suivi d'un numéro séquentiel (p. ex. blues001). Par conséquent, seul le genre de la pièce est connu. Les données utilisées pour bâtir cette collection proviennent de diverses sources (CD audio, radio, capture à l'aide de microphone, etc.) de manière à ce que les conditions variables d'enregistrement soient représentées. GTZAN est citée à de nombreuses reprises dans la littérature consacrée à la recherche de similarité musicale (Sturm, 2012). Sturm a répertorié 16 publications datant de 2004 à 2012 dont des résultats proviennent d'expérimentations effectuées sur cet ensemble de données. Ainsi GTZAN est considéré comme une référence dans le domaine du MIR. C'est la raison pour laquelle cette collection a été préférée à d'autres. Un échantillon de 200 pièces parmi les 1 000 contenues dans GTZAN a été sélectionné pour former l'ensemble témoin.

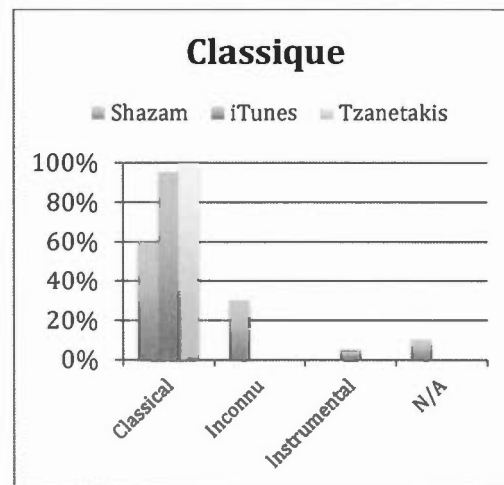
L'utilisation de GTZAN a été effectuée avec l'accord de Tzanetakis.

6.1.3. Résultats et discussion

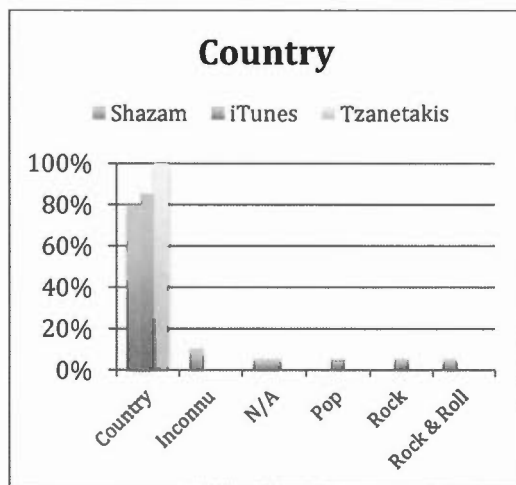
Selon l'hypothèse de Tzanetakis, les 200 pièces musicales de l'ensemble témoin sont présumées être réparties également au sein de 10 genres musicaux. L'évaluation des résultats obtenus a été faite en fonction de cette hypothèse. Au terme de l'expérimentation, 36 genres distincts ont été comptabilisés ce qui représente une augmentation de 360% par rapport à l'hypothèse de départ. Parmi les genres comptabilisés, seulement 6 (Blues, Classique, Country, Jazz, Pop et Reggae) ont été identifiés comme étant communs aux 3 schémas de classification. Le « Blues », le « Classique » et « Country » forment les genres les plus constants. Un genre est considéré comme étant constant lorsqu'il est ni sous-représenté ni surreprésenté, et ce pour les 3 modèles. Les genres « Pop » et « Rock » sont des genres surreprésentés tandis que les genres « Hip Hop », « Disco » et « Métal » sont des exemples de genre sous-représentés. Le fait qu'un genre soit surreprésenté ou sous-représenté indique l'absence de consensus par rapport à sa définition. Un constat important est que des termes différents sont utilisés pour décrire ce qui semble être le même genre musical. Par exemple, les termes « Rap » et « Hip Hop » sont utilisés pour représenter ce qui semble être le même genre musical. Même si ces associations peuvent paraître évidentes elles peuvent néanmoins représenter des classes distinctes dans certains modèles de classification. La sémantique des termes peut devenir un obstacle à l'exploration de données musicale lorsque la sémantique qu'un schéma attribue à un terme n'est pas la même que celle pressentie par celui qui désire explorer les données. Les figures 6.2 à 6.11 donnent le détail de la répartition des pièces en fonction du schéma de Tzanetakis.



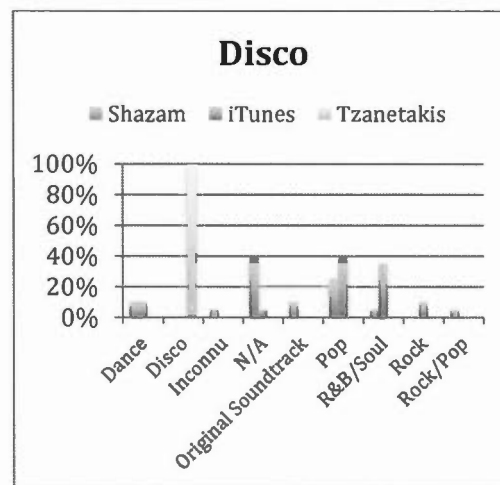
6.2. Répartition des pièces présumées
BLUES



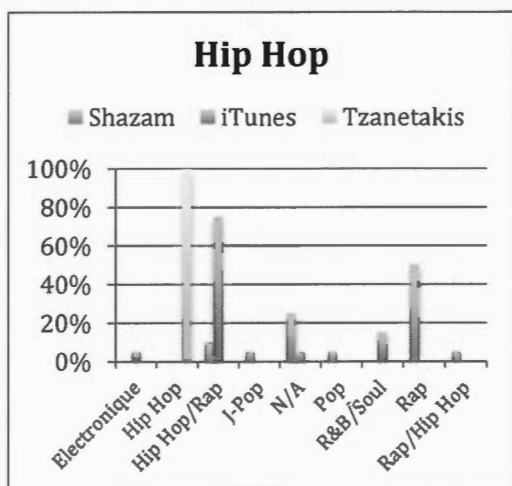
6.3. Répartition des pièces présumées
CLASSIQUE



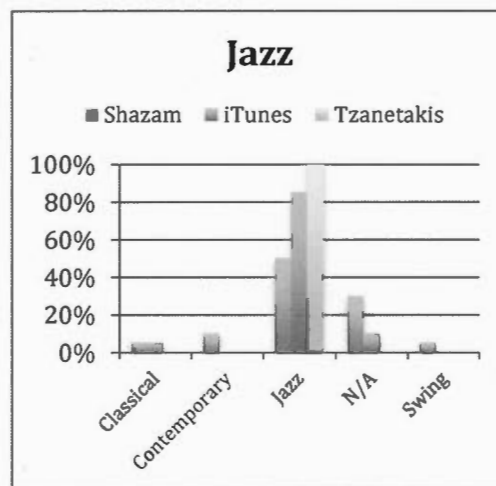
6.4. Répartition des pièces présumées
COUNTRY



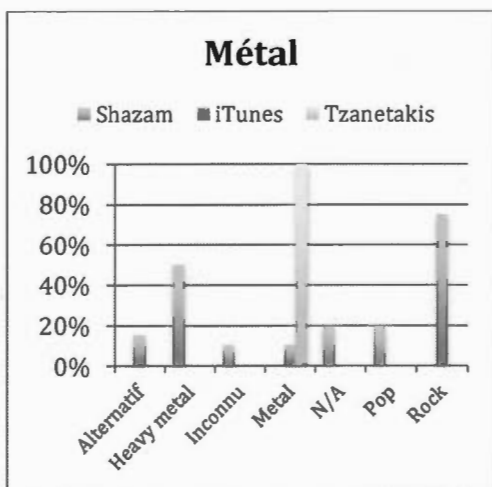
6.5. Répartition des pièces présumées
DISCO



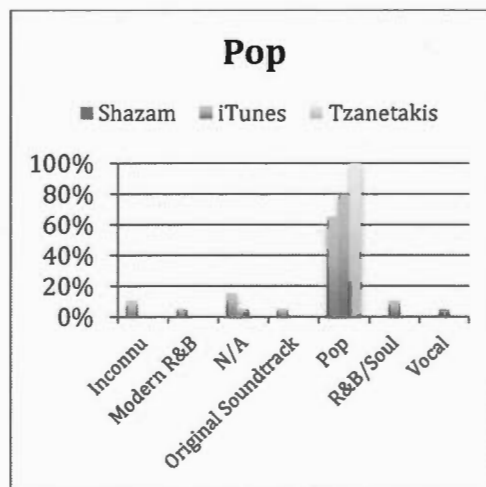
6.6. Répartition des pièces présumées HIP-HOP



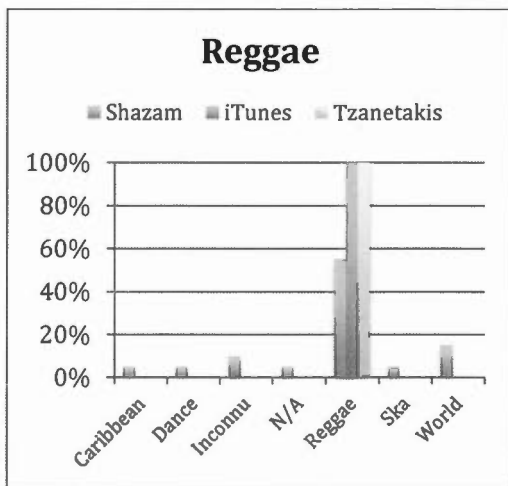
6.7. Répartition des pièces présumées JAZZ



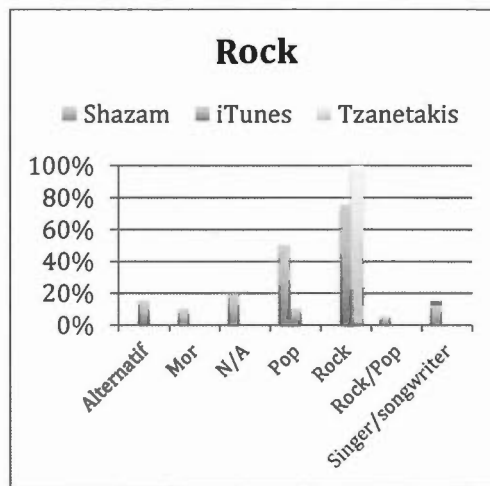
6.8. Répartition des pièces présumées MÉTAL



6.9. Répartition des pièces présumées POP



**6.10. Répartition des pièces présumées
REGGAE**



**6.11. Répartition des pièces présumées
ROCK**

Comme l'illustrent les figures 6.2 à 6.11, le modèle de Tzanetakis se rapproche davantage de celui d'iTunes que de celui de Shazam. Pourtant, Shazam est doté de mécanismes de reconnaissance de musique exceptionnels qui lui permettent de reconnaître la majorité des pièces musicales en quelques secondes (Wang, 2003). Les résultats obtenus lors de l'expérimentation I ont été évalués en fonction de schéma de Tzanetakis. Dans ce contexte, le modèle de classification de Shazam peut paraître bruité. Cependant, dans un contexte différent où l'ensemble témoin aurait reflété le schéma de Shazam alors l'interprétation des résultats aurait été tout autre. Il paraît donc convenable de prétendre que la qualité d'une classification dépend du contexte dans lequel cette classification est abordée.

Les résultats de l'expérimentation I soulignent le caractère subjectif du genre musical. Ces résultats renchérissent ceux de Rentfrow qui mènent à une certaine critique quant à l'utilisation du genre comme descripteur unique (Rentfrow et al., 2011). La classification d'un ensemble de musiques selon le genre est une tâche difficile à accomplir, et ce même pour un humain. La définition que l'on accorde à un genre détermine la nature des pièces susceptible d'y être associées. Les résultats obtenus lors de l'expérimentation I démontrent que ces définitions diffèrent d'un schéma de classification à l'autre. Bien qu'il existe des genres dont la définition semble être partagée, des étiquettes différentes peuvent être utilisées pour référer à ces genres. Cette observation est un autre facteur en défaveur du genre comme descripteur unique. C'est en partie en cause de ces limites qu'il est fréquent que plusieurs étiquettes soient utilisées pour représenter une pièce musicale. Les résultats de l'expérimentation I exposent des arguments en faveur de l'hypothèse selon laquelle l'exploration de données musicales peut être améliorée lorsque la musique est abordée selon différentes perspectives exprimées par des descripteurs appartenant de divers niveaux d'abstraction. En plus d'utiliser des étiquettes potentiellement subjectives, des descripteurs de bas niveau peuvent être inclus afin de préciser ce qui caractérise le contenu musical.

6.2. Expérimentation II : classification à l'aide de règles d'association

Comme il a été présenté au chapitre 3, la classification est une des stratégies les plus communes et des plus efficaces pour soutenir l'exploration de données musicales. Dans ce contexte, le fait qu'un système d'aide à l'exploration de données musicale soit doté de capacités classificatoires représente un avantage certain pour ce système. Basé sur cette affirmation, il convient donc d'évaluer la capacité des règles d'association pour servir la classification des données musicales. L'expérimentation II a été effectuée dans l'objectif de vérifier comment les règles d'association peuvent être utilisées pour soutenir la classification de données musicales à partir de descripteurs de bas niveau. Lors de cette expérimentation, un modèle de classification a été développé et testé. À l'instar de Arjannikov et Zhang, le modèle proposé utilise les motifs récurrents exprimés par les règles d'association pour mener la classification (Arjannikov et Zhang, 2014). Cependant, contrairement à ces derniers, l'approche proposée ne consiste pas à entraîner un classifieur pour obtenir des prototypes représentant des genres musicaux mais plutôt de comparer les associations extraites des pièces musicales afin de mesurer le degré de similarité structurelle des pièces. En plus d'évaluer la capacité des règles d'association à faciliter la mise en œuvre de mécanismes voués à l'exploration de données musicales à partir de descripteurs de bas niveau, cette expérimentation permet de vérifier l'hypothèse selon laquelle les relations de cooccurrence de certaines caractéristiques de l'onde sonore sont tout aussi importantes que la présence ou de l'absence de ces caractéristiques.

La musique peut être classifiée selon différents schémas. La période (p. ex. les années 60, le 19^e siècle, la période baroque, etc.), la localisation (Canada, É.-U., Amérique latine, Afrique, etc.) et l'émotion véhiculée (joie, tristesse, colère, etc.)

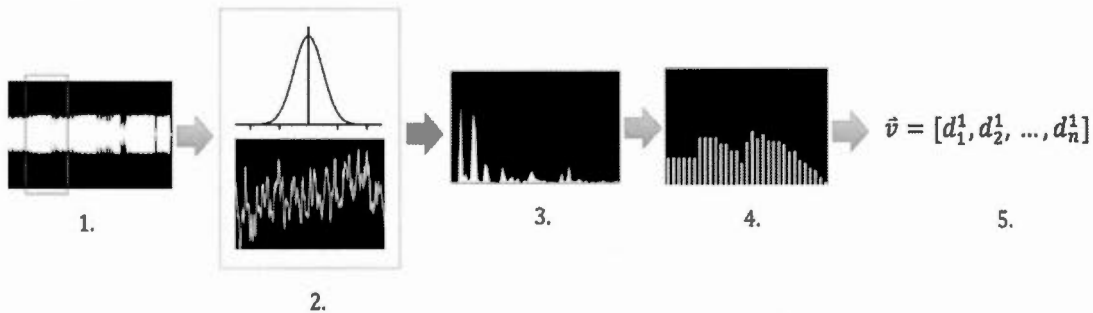
peuvent servir à bâtir ces schémas. Toutefois, l'appartenance à un genre musical demeure le critère le plus commun pour définir un schéma de classification. Pour cette raison, malgré sa subjectivité, le genre musical a été retenu comme critère.

6.2.1. Méthodologie

Afin d'être adaptées au processus de classification automatisée, les pièces musicales sont représentées par des vecteurs contenant des unités d'information qui traduisent les propriétés de l'onde sonore. La revue de la littérature exposée aux chapitres 2 et 3 révèle qu'il existe de nombreux mécanismes pour extraire des caractéristiques capables de représenter des propriétés de l'onde sonore qui sont suffisamment discriminantes pour être utilisées à des fins de classification. La sélection d'une caractéristique aux dépens d'un autre impacte les résultats produits. Quoiqu'il soit généralement admis que l'usage combiné d'une variété de caractéristiques donne de meilleurs résultats, une seule caractéristique a été retenue lors de cette expérimentation. Ce choix a été motivé par le désir de conserver la mise en œuvre de l'expérimentation aussi simple que possible. Si une classification intéressante est produite à l'aide d'une seule caractéristique alors les classifications produites avec plusieurs peuvent qu'être meilleure.

Comme il a été présenté au chapitre 5, l'utilisation des règles d'association est sujette à une problématique importante : les résultats qu'elles engendrent dépendent du nombre de variables considérées. Des données trop disparates peuvent entraîner une explosion combinatoire qui détériore les performances. Pour cette raison, un effort important a été consacré afin d'exercer un contrôle

sur la diversité des représentations vectorielles. La figure 6.12 illustre le processus établi.



6.12. Processus de la création de la représentation vectorielle des fichiers

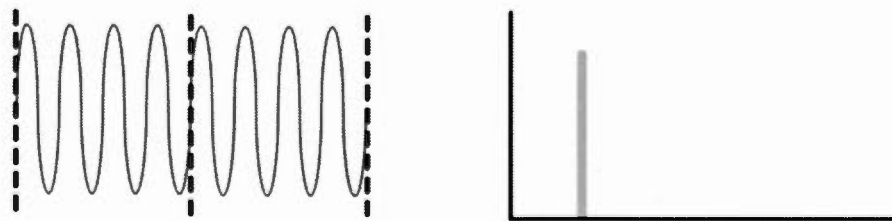
Le processus de création des représentations vectorielle utilisée pour exprimer les propriétés de l'onde sonore est composé de 5 étapes :

- 1) La segmentation des données ;
- 2) L'application d'une fonction de fenêtrage ;
- 3) L'application d'une transformée de Fourier ;
- 4) Le regroupement fréquentiel ;
- 5) La sélection des bandes de fréquence jugées les plus descriptives.

La première étape du processus consiste à segmenter les données de manière à obtenir des segments d'une durée approximative de 46 millisecondes (2048 points d'échantillonnage). Cette opération génère des échantillons dont le rôle

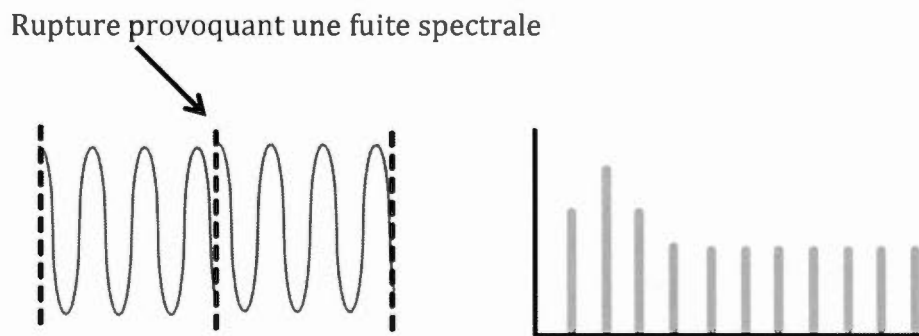
est de décrire l'évolution du signal dans le temps. La taille des segments résulte d'un compromis entre précision fréquentielle et temporelle. En effet, plus la taille de la fenêtre est grande, plus précise est la détection des fréquences. Cependant, plus la taille de la fenêtre est grande, plus il devient difficile de détecter exactement le moment où un événement sonore se manifeste dans le temps.

La représentation vectorielle des données est créée à partir d'une estimation de l'empreinte fréquentielle des segments. Pour obtenir cette empreinte, une transformée de Fourier est appliquée. La seconde étape du processus de création des représentations vectorielle consiste à exécuter une fonction de fenêtrage sur les données segmentées afin de limiter la fuite spectrale. La fuite spectrale est un phénomène qui se produit lorsque l'intervalle temporel du signal soumis à une FFT ne comprend pas un nombre entier de périodes. L'une des particularités de la FFT est qu'elle considère que le signal analysé est circulaire c'est-à-dire que ses deux extrémités peuvent être reliées l'une à l'autre. Lorsque le signal représenté est parfaitement périodique et que l'intervalle temporel considéré comprend un nombre entier de périodes, la FFT génère une représentation fidèle du signal (figure 6.13).



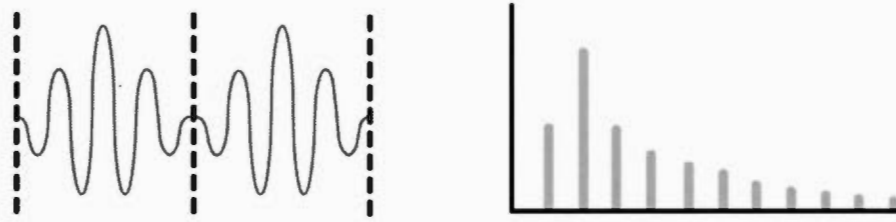
6.13. FFT appliquée à un signal périodique

La portion du signal contenu dans les différents segments possède rarement la propriété requise pour que la FFT fonctionne parfaitement. Les segments contiennent généralement des signaux non circulaires. Cette discontinuité a comme effet d'incorporer des fréquences inattendues dans le signal. C'est ce qu'on appelle la fuite spectrale. Ainsi, les signaux discontinus contenus dans les segments ne représentent pas fidèlement le signal d'origine, car ils possèdent des spectres fréquentiels différents (figure 6.14).



6.14. FFT appliquée à un signal non périodique

L'utilisation d'une fonction de fenêtrage permet de réduire la distorsion créée par la présence de signaux tronqués. Une telle fonction opère dans le domaine temporel. Elle adapte le signal de manière à ce que ses extrémités puissent être reliées. Les transitions brusques sont ainsi éliminées ce qui offre une meilleure résolution du spectre (figure 6.15).



6.15. FFT appliquée à un signal filtré

Il existe plusieurs fonctions de fenêtrage (Blackman, Hamming, Hann, etc.). Ces fenêtres offrent différentes résolutions fréquentielles / amplitudes. Dans le cadre de cette expérimentation, la fenêtre de Hamming a été utilisée, car elle favorise la résolution fréquentielle. La forme de la fenêtre de Hamming est donnée par l'équation 6.1 où n et N représentent respectivement le n ème point d'échantillonnage et la taille du segment.

$$w_{\text{hamming}}(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N}\right), 0 \leq n \leq N \quad (6.1)$$

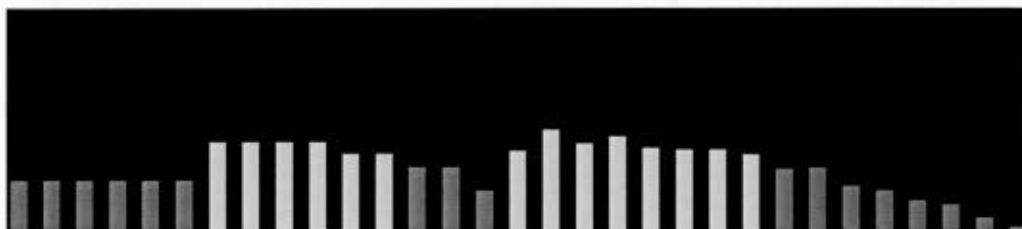
La troisième étape du processus de création des représentations vectorielles consiste à appliquer la transformée de Fourier aux segments préalablement ajustés à l'aide de la fonction de fenêtrage de manière à obtenir le spectre fréquentiel de chacun des segments. La librairie *Accelerate Framework* a été utilisée pour ce faire. Le spectre fréquentiel obtenu suite à l'application de la FFT est ensuite subdivisé en bandes de fréquences.

Le découpage des fréquences en bandes correspond à la cinquième étape du processus de création des représentations vectorielle. Cette opération vise à maintenir à un niveau acceptable le nombre d'items qui seront contenus dans les transactions. Ainsi, avec un découpage en 31 bandes, le nombre maximum d'items contenus dans une transaction est fixé à 31 soit exactement le nombre de bandes. Deux schémas sont proposés, un découpage en 9 bandes et un découpage en 31 bandes. L'usage du découpage en 9 bandes permet de réduire le nombre de calculs nécessaire à l'extraction des règles d'association. Cependant, ce découpage engendre une forme de compression qui peut entraîner une perte d'information. Pour cette raison, le découpage en 31 bandes a été privilégié. Le schéma utilisé pour grouper les bandes de fréquence en 31 bandes est donné au tableau 6.1.

Bande de fréquence	Fréquence minimum (Hz)	Fréquence maximum (Hz)	Bande de fréquence	Fréquence minimum (Hz)	Fréquence maximum (Hz)
1	0	21	17	625	796
2	22	43	18	797	1 012
3	44	64	19	1 013	1 248
4	65	86	20	1 249	1 614
5	87	107	21	1 615	2 002
6	108	129	22	2 003	2 497
7	130	150	23	2 498	3 143
8	151	172	24	3 144	4 005
9	173	193	25	4 006	4 996
10	194	214	26	4 996	6 309
11	215	236	27	6 310	8 010
12	237	258	28	8 011	9 991
13	259	322	29	9 992	12 510
14	323	409	30	12 511	15 999
15	410	516	31	16 000	22 050
16	517	624			

6.1. Découpage du spectre fréquentiel en bande de fréquence

La dernière étape du processus de création des représentations vectorielles consiste à retenir les fréquences les plus importantes. Une bande de fréquence est jugée significative si sa contribution à l'énergie du segment est supérieure à un seuil prédéterminé. Ce seuil est appelé seuil de participation et est noté p . La figure 6.16 représente cette sélection. Les bandes surlignées en rouge sont jugées non significatives et par conséquent elles sont rejetées.

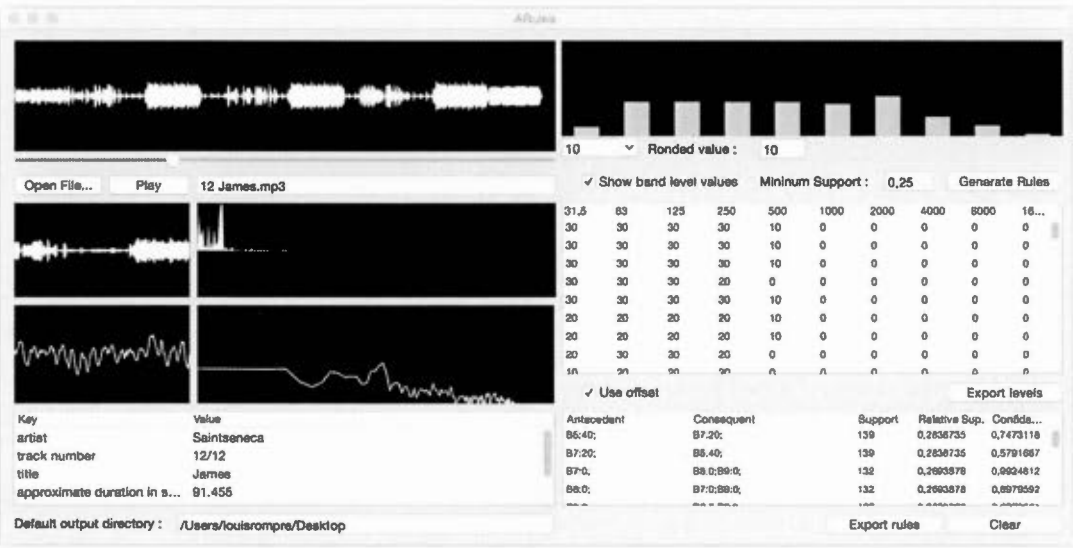


6.16. Élimination des bandes de fréquences jugées non significatives

Les bandes illustrées en rouge sont éliminées.

Selon la méthodologie proposée, une pièce musicale de 3 minutes est exprimée par 7 826 représentations vectorielles. Ces représentations vectorielles sont considérées comme étant les transactions. C'est donc à partir de ces représentations vectorielles que les règles d'associations sont extraites. L'extraction des règles d'association est réalisée à l'aide de l'algorithme APRIORI présenté au chapitre 4. Cette opération est la plus coûteuse du processus. Les règles d'associations obtenues expriment des relations de cooccurrence entre certaines propriétés de l'onde sonore. C'est à partir d'une évaluation de ces relations que le degré de similarité entre deux pièces musicales est établi. La

figure 6.17 illustre l'interface permettant de définir la valeur des différents paramètres.



6.17. Capture d'écran du module d'extraction des descripteurs de bas niveau

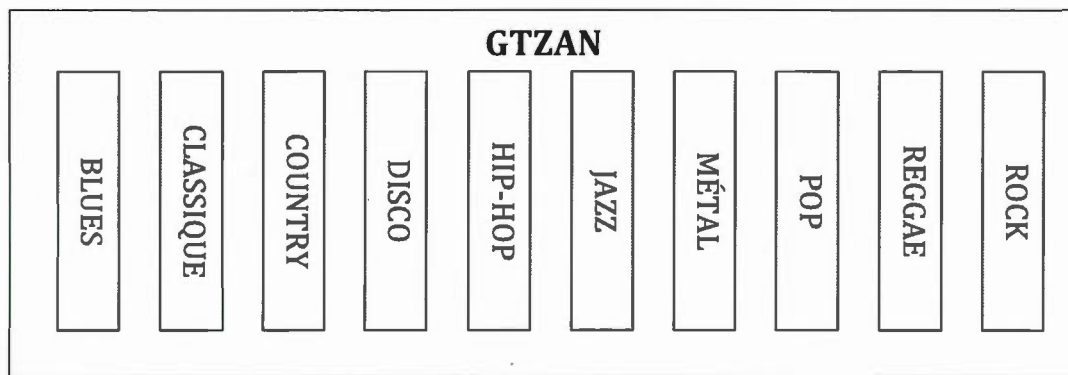
6.2.2. Ensemble de données

Tout comme lors de l'expérimentation I, l'ensemble de données utilisé lors de cette expérimentation est GTZAN. Cet ensemble de données a été sélectionné afin qu'il soit possible de comparer les résultats obtenus avec ceux obtenus lors de l'expérimentation I. À titre de rappel, l'ensemble GTZAN est composé de 1 000 pièces musicales classifiées en fonction de 10 genres musicaux. Chacun des genres est représenté par 100 pièces. Les pièces musicales sont numérotées pour être identifiées. Comme démontré à l'expérimentation I, le modèle de

classification adopté pour classier l'ensemble GTZAN représente le point de vue de Tzanetakis et son évaluation demeure subjective. L'objectif de l'expérimentation n'est donc pas de chercher à reproduire à tout prix le modèle de Tzanetakis mais plutôt d'utiliser les règles d'association afin de découvrir des relations cachées pouvant être utilisées pour servir le processus de classification.

6.2.3. Résultats et discussion

Plusieurs résultats ont été comptabilisés lors de l'expérimentation II. Dans un premier temps, la capacité de la solution proposée à classier les données a été évaluée. Pour ce faire, les classes de similarité définies dans GTZAN ont été utilisées comme repère.



6.18. Schéma utilisé comme repère lors de l'expérimentation II

Le schéma de GTZAN a été utilisé comme repère. Les genres musicaux forment des classes de référence.

Les règles d'association ont été extraites pour la totalité de documents audio contenus dans GTZAN. Les tableaux 6.2 et 6.3 donnent respectivement le nombre de règles d'association extraites pour 20 pièces musicales présumées classiques et 20 pièces musicales présumées Rock.

Classique	
ID	Règles produites
95	506
90	714
85	3026
80	664
75	198
70	260
65	126
60	1824
55	76
50	810
45	310
40	400
35	234
30	478
25	812
20	660
15	246
10	212
5	340
0	254

Moyenne : **607,5**

6.2. Règles extraites pour un sous ensemble présumé classique

Rock	
ID	Nombre de règles
95	512
90	98
85	132
80	250
75	290
70	420
65	368
60	618
55	2202
50	530
45	850
40	136
35	274
30	222
25	1276
20	2046
15	156
10	156
5	3348
0	420

Moyenne : **715,2**

6.3. Règles extraites pour un sous ensemble présumé rock

L'examen des données des tableaux 6.2 et 6.3 révèle que pour un même genre présumé le nombre de règles d'association extraites varie considérablement d'une pièce musicale à l'autre. Afin de gérer cet écart, les règles d'association ont été triées en ordre décroissant en fonction de leur support et seulement les x premières règles ont été retenues lors de l'évaluation. Plusieurs essais ont été effectués afin de cibler le bon nombre de règles à considérer. Lors de l'exécution des expérimentations, il a été constaté que considérer 15 règles était suffisant pour produire des résultats intéressants.

Afin d'être en mesure d'établir un niveau d'appartenance d'une pièce à un genre, un sous-ensemble de pièces a été sélectionné au hasard pour représenter les différents genres. Les règles d'association des pièces à classifier ont été comparées à celles des pièces sélectionnées pour représenter les genres. Un degré de similarité est calculé pour chacun des genres représentés.

GENRE	EFFICACITÉ
BLUES	25%
CLASSIQUE	90%
COUNTRY	20%
DISCO	15%
HIP-HOP	10%
JAZZ	15%
MÉTAL	60%
POP	40%
REGGAE	50%
ROCK	5%

6.4. Pourcentage de cohérence avec GTZAN

Le tableau 6.4 donne, pour chacun des genres, le pourcentage de documents dont le degré de similarité le plus élevé renvoie au genre auquel ils sont présumés appartenir. Ces résultats ont été obtenus avec un support fixé à 10% et un seuil de participation fixé à 40%. On remarque que pour certains genres, la solution proposée est en mesure de bien traduire le schéma de classification de GTZAN. C'est le cas notamment pour les pièces présumées classique. 90% des pièces présumées classiques ont été classifiées en accord avec le modèle de référence. Ce score rivalise avec les scores exposés dans la littérature lorsque les propriétés de l'onde sonore sont considérées pour reconnaître et classifier la musique classique (Tzanetakis, 2002 ; Sigtia et Dixon, 2014, Baniya et al., 2014). Le tableau 6.5 donne le niveau d'appartenance aux différents genres de 20 pièces présumées classique. L'examen des données du tableau 6.5 révèle que le degré de similarité de toutes les pièces présumées classiques à l'exception des pièces CL95²⁶ et CL90 est nettement plus élevé pour le genre classique que n'importe quel autre genre. Ce résultat collabore ceux de Tzanetakis (Tzanetakis et Cook, 2002) et (Deshpande et al., 2001) qui suggèrent que le genre classique est le genre musical le plus propice au processus de classification automatisée. Une autre observation intéressante est que les pièces CL95 et CL90 sont toutes deux jugées plus similaire au schéma correspondant au genre métal qu'au genre classique. Quoique cette observation puisse sembler bruitée à l'écoute des pièces, certaines études suggèrent qu'il existe une relation entre la musique classique et le métal (North et Hargreaves, 2008). North et Hargreaves ont mené une étude sur les préférences musicales et leur lien probable avec des traits de personnalité. Les résultats de leur recherche démontrent qu'il existe des similitudes étonnantes entre les admirateurs de musique classique et métal. De plus, ces deux genres musicaux partagent certaines similitudes structurales. Les

²⁶ L'identifiant classical.00095 initialement attribué par Tzanetakis a été remplacé par CL95 afin qu'il soit plus compact.

pièces classiques et métal ont tendance à mettre en valeur des arrangements sonores complexes. Ainsi, dans un contexte d'exploration de données musicale il peut être pertinent de joindre à la fois des pièces musicales classique et métal.

Le pourcentage global de précision obtenu lors de l'expérimentation II est grandement impacté par le faible score accordé à certains genres tels que le rock et le hip-hop. Parmi tous les genres considérés, le rock est celui qui est le plus problématique. Le tableau 6.6 donne le niveau d'appartenance aux différents genres de 20 pièces présumées rock. L'analyse des données du tableau 6.6 ne permet pas de cibler une corrélation significative entre les pièces musicales présumées rock et un genre particulier. Les motifs récurrents exprimés par les règles d'association dégagées des pièces musicales présumées rock ne semblent pas être assez discriminants pour permettre au système de distinguer le rock des autres genres musicaux. La difficulté de classifier de la musique rock à partir de traitements automatisés a déjà été soulevée par Tzanetakis (Tzanetakis et Cook, 2002) et Deshpande (Deshpande et al., 2001) et Baniya (Baniya et al., 2014).

Classique

ID	Rock	REGGAE	POPULAIRE	MÉTAL	JAZZ	HIP-HOP	DISCO	COUNTRY	CLASSIQUE	BLUES	MAX
CL95	20,00%	4,78%	1,49%	30,15%	8,66%	6,87%	13,43%	2,09%	4,48%	8,06%	30,15%
CL90	16,32%	3,26%	6,06%	20,28%	4,66%	6,29%	19,58%	3,03%	16,55%	3,96%	20,28%
CL85	8,76%	3,37%	8,76%	2,16%	11,99%	4,99%	13,75%	8,09%	29,92%	8,22%	29,92%
CL80	7,57%	3,89%	9,00%	0,20%	19,02%	2,86%	6,54%	18,20%	21,68%	11,04%	21,68%
CL75	4,76%	4,02%	6,24%	0,85%	22,83%	4,23%	2,85%	12,47%	34,57%	7,19%	34,57%
CL70	6,07%	4,13%	5,34%	2,91%	15,29%	5,10%	6,43%	11,65%	34,47%	8,62%	34,47%
CL65	5,75%	4,66%	8,77%	0,96%	20,82%	6,16%	7,26%	12,47%	23,70%	9,45%	23,70%
CL60	7,99%	3,03%	8,26%	2,20%	15,70%	3,86%	9,09%	10,28%	32,23%	7,35%	32,23%
CL55	6,15%	4,07%	9,05%	0,83%	21,35%	4,65%	4,24%	13,54%	28,57%	7,56%	28,57%
CL50	5,68%	4,41%	6,76%	0,54%	22,34%	4,86%	4,32%	13,96%	29,28%	7,84%	29,28%
CL45	6,87%	3,44%	10,80%	1,88%	16,78%	4,42%	8,27%	12,36%	26,43%	8,76%	26,43%
CL40	7,31%	3,89%	13,11%	2,47%	17,09%	4,56%	9,02%	11,59%	22,70%	8,26%	22,70%
CL35	9,19%	5,14%	6,92%	1,19%	18,28%	3,66%	5,04%	17,89%	20,55%	12,15%	20,55%
CL30	5,46%	3,66%	7,86%	0,73%	20,77%	5,66%	6,13%	13,25%	28,96%	7,52%	28,96%
CL25	4,11%	2,66%	3,50%	0,36%	17,63%	5,80%	4,11%	14,86%	40,22%	6,76%	40,22%
CL20	4,34%	3,30%	6,60%	0,87%	20,57%	4,86%	3,30%	12,24%	36,72%	7,20%	36,72%
CL15	14,43%	3,48%	6,47%	2,99%	6,97%	1,00%	17,91%	4,98%	28,36%	13,43%	28,36%
CL10	5,68%	3,41%	5,68%	1,57%	15,20%	5,68%	7,16%	12,66%	35,55%	7,42%	35,55%
CL05	5,80%	3,01%	8,96%	0,90%	19,05%	4,82%	6,70%	13,48%	30,35%	6,93%	30,35%
CL00	8,27%	4,74%	6,91%	2,98%	14,09%	3,52%	12,06%	6,23%	34,55%	6,64%	34,55%
	8,03%	3,82%	7,33%	3,85%	16,45%	4,69%	8,36%	11,27%	27,99%	8,22%	90 %

6.5. Degré de similarité de 20 pièces musicales présumées CLASSIQUE

Rock

ID	Rock	REGGAE	POPULAIRE	MÉTAL	JAZZ	HIP-HOP	DISCO	COUNTRY	CLASSIQUE	BLUES	MAX
R95	12,47%	6,45%	6,88%	5,38%	16,13%	6,24%	7,53%	13,33%	12,47%	13,12%	16,13%
R90	9,03%	6,11%	9,44%	4,03%	9,58%	4,03%	13,33%	8,89%	28,06%	7,50%	28,06%
R85	10,31%	8,02%	11,63%	3,17%	12,42%	5,46%	14,45%	10,22%	15,07%	9,25%	15,07%
R80	7,53%	3,21%	13,13%	1,38%	18,82%	3,95%	8,17%	14,78%	19,93%	9,09%	19,93%
R75	10,40%	13,20%	11,84%	9,67%	6,24%	17,00%	11,12%	9,67%	0,54%	10,31%	17,00%
R70	11,06%	6,98%	6,89%	3,30%	16,39%	7,76%	7,47%	16,39%	10,67%	13,09%	16,39%
R65	17,65%	6,51%	4,91%	17,65%	6,54%	8,18%	17,65%	5,26%	9,82%	5,84%	17,65%
R60	12,14%	5,11%	8,31%	3,83%	13,42%	1,76%	17,41%	5,43%	25,56%	7,03%	25,56%
R55	4,53%	2,40%	6,38%	0,74%	17,10%	6,38%	6,56%	12,75%	36,04%	7,12%	36,04%
R50	4,58%	18,74%	23,44%	5,28%	5,82%	16,38%	8,30%	8,46%	1,40%	7,61%	23,44%
R45	10,55%	4,75%	5,94%	3,42%	14,86%	4,90%	6,39%	21,25%	10,25%	17,68%	21,25%
R40	12,55%	12,96%	5,94%	18,76%	3,64%	12,42%	12,82%	10,26%	0,54%	10,12%	18,76%
R35	6,79%	15,64%	18,72%	8,72%	5,00%	16,67%	9,36%	7,44%	0,77%	10,90%	18,72%
R30	6,39%	15,85%	11,57%	6,83%	10,24%	10,68%	10,02%	12,56%	5,51%	10,35%	15,85%
R25	16,61%	4,44%	2,63%	42,27%	2,96%	11,02%	13,49%	1,81%	1,48%	3,29%	42,27%
R20	8,20%	5,76%	9,47%	2,64%	17,87%	5,08%	8,20%	12,40%	20,31%	10,06%	20,31%
R15	5,08%	11,67%	10,85%	3,64%	15,94%	10,17%	5,69%	13,05%	13,74%	10,17%	15,94%
R10	15,95%	8,45%	6,22%	5,10%	8,93%	10,37%	9,09%	13,88%	2,07%	19,94%	19,94%
R05	2,46%	3,29%	2,87%	0,82%	15,20%	7,39%	4,93%	9,45%	44,56%	9,03%	44,56%
R00	9,81%	9,02%	6,90%	4,38%	13,40%	5,17%	7,82%	17,37%	14,46%	11,67%	17,37%

9,70%	8,43%	9,20%	7,55%	11,53%	8,55%	9,99%	11,23%	13,66%	10,16%	5 %
-------	-------	-------	-------	--------	-------	-------	--------	--------	--------	-----

6.6. Degré de similarité de 20 pièces musicales présumées ROCK

En moyenne, la solution proposée est en mesure de reproduire 33 % du schéma de classification initial de GTZAN, et ce sans avoir été préalablement entraîné pour exécuter cette tâche. Même si ce pourcentage est inférieur à d'autres déjà publiés dans la littérature, il demeure néanmoins intéressant pour une approche non supervisée utilisant une seule unité d'information pour traduire l'onde sonore. En effet, la majorité des approches proposées pour classier automatiquement l'ensemble de données GTZAN repose sur l'usage de systèmes de classification à apprentissage supervisés opérant sur plusieurs caractéristiques. Les résultats de l'expérimentation II permettent de conclure que l'usage des règles d'association dans un contexte de classification de données sonore, quoique non optimisé, est une avenue intéressante. Il est probable que les résultats produits puissent être améliorés en mixant plusieurs caractéristiques extraites de l'onde sonore comme c'est le cas pour les systèmes les plus performants (Baniya et al., 2014).

La capacité de l'approche proposée à bien classier la musique classique et sa difficulté à classier la musique rock sont des caractéristiques observées chez d'autres systèmes spécialisés dans la classification de données sonores. L'avantage des règles d'association est qu'elles peuvent être utilisées afin de pouvoir identifier quels sont les descripteurs à l'origine de ces résultats. Les tableaux 6.8 et 6.9 donnent les 20 premières règles d'association des deux pièces qui ont le degré de similarité envers le genre classique le plus élevé et les 20 premières règles d'association des deux pièces qui ont le degré de similarité envers le genre rock le plus élevé.

CL20			
Antécédent	Conséquent	Sup.	Conf.
1013;	517;	0,5	0,8
517;	1013;	0,5	0,6
323;	517;	0,4	0,8
517;	323;	0,4	0,6
410;	517;	0,4	0,8
517;	410;	0,4	0,6
625;	517;	0,4	0,7
517;	625;	0,4	0,6
797;	517;	0,4	0,7
517;	797;	0,4	0,5
1013;	323;	0,3	0,6
323;	1013;	0,3	0,6
323;	625;	0,3	0,6
625;	323;	0,3	0,6
410;	625;	0,3	0,6
625;	410;	0,3	0,6
410;	323;	0,3	0,6
323;	410;	0,3	0,5
797;	410;	0,3	0,6
410;	797;	0,3	0,6

CL25			
Antécédent	Conséquent	Sup.	Conf.
410;	517;	0,6	0,8
517;	410;	0,6	0,7
625;	517;	0,5	0,8
517;	625;	0,5	0,7
625;	410;	0,5	0,8
410;	625;	0,5	0,7
410;625;	517;	0,4	0,8
517;625;	410;	0,4	0,8
410;517;	625;	0,4	0,7
625;	410;517;	0,4	0,6
410;	517;625;	0,4	0,6
517;	410;625;	0,4	0,5
797;	517;	0,4	0,8
517;	797;	0,4	0,5
323;	517;	0,4	0,8
517;	323;	0,4	0,5
1249;	410;	0,3	0,8
410;	1249;	0,3	0,5
797;	410;	0,3	0,8
410;	797;	0,3	0,5

6.8. Règles d'association de deux pièces présumées CLASSIQUE

Les règles d'association listées dans le tableau 6.8 appartiennent aux deux pièces musicales dont le degré de similarité envers le classique est le plus élevé. Le degré de similarité de la pièce CL20 envers le classique est de 36,72% tandis que le degré de similarité de la pièce CL25 envers le genre classique est de 40,22%. Le tableau 6.8 liste seulement les 20 premières règles des deux pièces.

R25			
Antécédent	Conséquent	Sup.	Conf.
3144;	4006;	0,4	0,8
4006;	3144;	0,4	0,6
2498;	4006;	0,4	0,9
4006;	2498;	0,4	0,5
2003;	4006;	0,4	0,9
4006;	2003;	0,4	0,5
797;	4006;	0,4	0,8
4006;	797;	0,4	0,5
2498;	3144;	0,3	0,8
3144;	2498;	0,3	0,7
1615;	4006;	0,3	0,8
4006;	1615;	0,3	0,5
2003;	3144;	0,3	0,8
3144;	2003;	0,3	0,6
2498;3144;	4006;	0,3	0,9
2498;4006;	3144;	0,3	0,8
3144;4006;	2498;	0,3	0,7
2498;	3144;4006;	0,3	0,7
3144;	2498;4006;	0,3	0,6
4006;	2498;3144;	0,3	0,4

R65			
Antécédent	Conséquent	Sup.	Conf.
1013;	797;	0,3	0,6
797;	1013;	0,3	0,6
2498;	3144;	0,3	0,7
3144;	2498;	0,3	0,6
2003;	3144;	0,3	0,7
3144;	2003;	0,3	0,6
1615;	797;	0,2	0,6
797;	1615;	0,2	0,5
2498;	2003;	0,2	0,7
2003;	2498;	0,2	0,6
1615;	1013;	0,2	0,6
1013;	1615;	0,2	0,5
1615;	3144;	0,2	0,6
3144;	1013;	0,2	0,5
3144;	1615;	0,2	0,5
1013;	3144;	0,2	0,5
3144;	797;	0,2	0,5
797;	3144;	0,2	0,5
130;	151;	0,2	0,5
151;	130;	0,2	0,5

6.9. Règles d'association de deux pièces présumées ROCK

Les règles d'association listées dans le tableau 6.9 appartiennent aux deux pièces musicales dont le degré de similarité envers le genre ROCK est le plus élevé. Le degré de similarité de la pièce R25 envers le genre ROCK est de 16,61% tandis que le degré de similarité de la pièce R65 envers le genre ROCK est de 17,65%. Le tableau 6.9 liste seulement les 20 premières règles des deux pièces.

R25			
Antécédent	Conséquent	Sup.	Conf.
3144;	4006;	0,4	0,8
4006;	3144;	0,4	0,6
2498;	4006;	0,4	0,9
4006;	2498;	0,4	0,5
2003;	4006;	0,4	0,9
4006;	2003;	0,4	0,5
797;	4006;	0,4	0,8
4006;	797;	0,4	0,5
2498;	3144;	0,3	0,8
3144;	2498;	0,3	0,7
1615;	4006;	0,3	0,8
4006;	1615;	0,3	0,5
2003;	3144;	0,3	0,8
3144;	2003;	0,3	0,6
2498;3144;	4006;	0,3	0,9
2498;4006;	3144;	0,3	0,8
3144;4006;	2498;	0,3	0,7
2498;	3144;4006;	0,3	0,7
3144;	2498;4006;	0,3	0,6
4006;	2498;3144;	0,3	0,4

M65			
Antécédent	Conséquent	Sup.	Conf.
2498;	3144;	0,6	0,9
3144;	2498;	0,6	0,7
4006;	3144;	0,6	0,9
3144;	4006;	0,6	0,7
4006;	2498;	0,5	0,7
2498;	4006;	0,5	0,7
2498;4006;	3144;	0,4	0,9
3144;4006;	2498;	0,4	0,8
2498;3144;	4006;	0,4	0,8
4006;	2498;3144;	0,4	0,7
2498;	3144;4006;	0,4	0,6
3144;	2498;4006;	0,4	0,6
172;	3144;	0,4	0,8
3144;	172;	0,4	0,6
2003;	3144;	0,4	0,9
3144;	2003;	0,4	0,5
237;	3144;	0,4	0,8
3144;	237;	0,4	0,5
237;	172;	0,4	0,8
172;	237;	0,4	0,7

6.10. Règles d'association de deux pièces présumées ROCK et MÉTAL

Même si les deux pièces sont présumées appartenir à des genres différents, elles partagent un nombre important de règles d'association.

En examinant les données des tableaux 6.8 on constate que 60% des 20 premières règles d'association sont les même pour les deux pièces ayant le degré de similarité le plus élevé envers le genre classique. Le nombre de règles partagées par les deux pièces rock est moindre. Le tableau 6.9 indique que seulement 20% des 20 premières règles d'associations sont partagées par les deux pièces ayant le degré de similarité le plus élevé envers le genre rock. Ce pourcentage explique le faible degré de similarité accordé à ces pièces envers le genre rock. En comparant les règles d'association des pièces présumées appartenir au genre rock avec celles des pièces qui définissent les autres genres, on constate que ces pièces partagent plusieurs règles d'association. En fait, le nombre de règles d'association partagées entre une pièce présumée appartenir au genre rock et une pièce présumée appartenir au genre métal peut atteindre jusqu'à 65% (tableau 6.10).

L'expérimentation II a mené à des résultats intéressants. Les résultats obtenus démontrent que les règles d'association sont en mesure de manipuler des descripteurs de bas niveau afin de soutenir le processus de classification. Ces résultats collaborent ceux de Conklin, Arjannikov et Zhang (Conklin, 2009; Arjannikov et Zhang, 2014). Ils suggèrent que les relations de cooccurrence exprimées par les règles d'association sont informatives et discriminantes. Une première analyse des résultats obtenus a permis de faire ressortir que le classique est le genre qui est le plus propice à la démarche tandis que le rock est celui qui s'y apprête le moins. Ce qui accorde une valeur ajoutée aux règles d'association c'est qu'elles permettent de mesurer l'impact des descripteurs sur les classifications générées. Le fait que les règles d'association ne soient pas opaques a permis de faire une analyse plus approfondie et ainsi découvrir quels étaient les motifs à l'origine des résultats obtenus. La découverte de motifs discriminants offre de nombreuses possibilités. Par exemple, ils peuvent être

stockés dans une base de données et utilisés comme critères de recherche pour accéder à des pièces musicales jugées similaires (consulter l'annexe A pour une démonstration). L'approche proposée bonifie la recherche de similarité musicale, car elle permet de cibler les descripteurs utiles en plus de préciser dans quels contextes ces descripteurs peuvent servir. Puisqu'il n'existe pas de théorie formelle pour guider la sélection de l'unité d'information à privilégier pour soutenir le processus de classification, cette faculté est des plus intéressantes.

6.3. Expérimentation III : analyse de la structure de pièces musicales

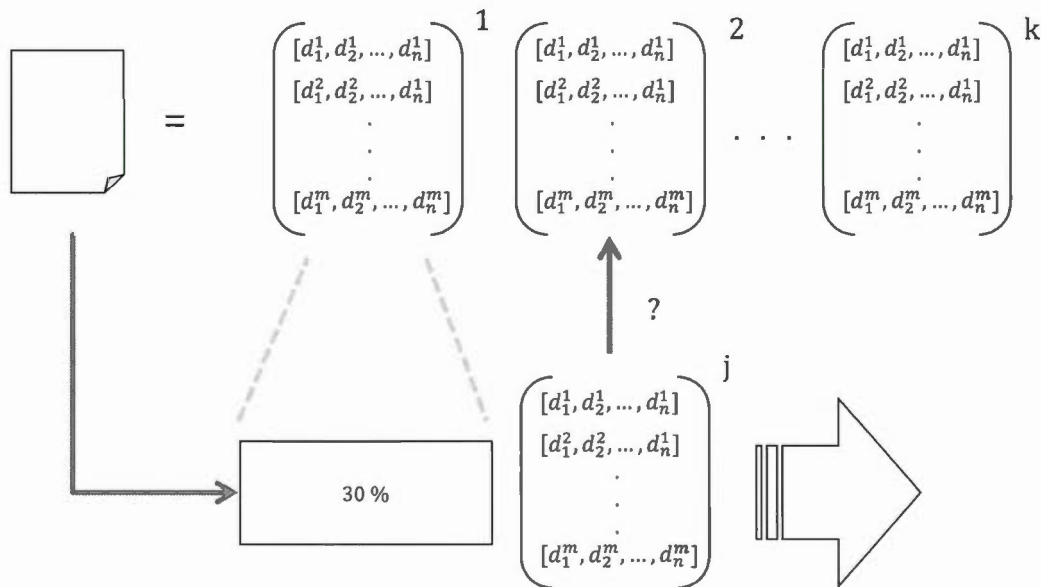
L'une des hypothèses soutenues dans cette thèse est que l'utilisation des règles d'association peut d'une part servir à traiter des informations provenant de différents niveaux d'abstraction et d'autre part à faciliter la mise en œuvre de différentes stratégies vouées à l'exploration de données musicales. C'est principalement la deuxième partie de cette hypothèse que cette expérimentation cherche à vérifier. L'exploration de données musicales peut se faire au sein d'un ensemble de documents audio mais également à l'intérieur d'un même document. La musique est une organisation séquentielle de sons qui se manifeste dans le temps. La manière dont les sons sont agencés forme ce qu'on appelle la structure musicale. L'analyse de la structure musicale est un des sujets d'étude de la musicologie. La démarche consiste à décomposer et à manipuler les pièces musicales de diverses manières afin de dégager des éléments et des règles qui dictent la façon dont ces éléments sont combinés (Ruwet, 1966). L'expérimentation III a été effectuée afin de vérifier s'il était possible d'utiliser les règles d'association pour effectuer une analyse de la structure musicale de

différentes pièces. Pour ce faire, une démarche similaire à celle adoptée lors de l'expérimentation II a été utilisée. Cependant, plutôt que de comparer les représentations vectorielles d'un ensemble de pièces avec celles d'autres pièces musicales, les représentations vectorielles utilisées pour représenter une pièce musicale ont été comparées entre elles afin de faire ressortir des motifs cachés à l'intérieur des pièces musicales. L'expérimentation III vise à vérifier la portabilité des règles d'association.

6.3.1. Méthodologie

La distinction entre la méthodologie adoptée lors de cette expérimentation et celle employée lors de l'expérimentation II réside d'une part dans l'ajout d'une étape de segmentation des pièces musicales et d'autre part dans le fait que les pièces musicales ont été comparées avec elles-mêmes plutôt qu'avec d'autres. C'est à partir d'une analyse comparative des segments qui la composent que la structure d'une pièce musicale est dégagée.

Les pièces musicales sont segmentées afin qu'il soit possible de décortiquer l'évolution du signal sonore dans le temps. L'analyse comparative est réalisée en deux étapes. La première étape consiste à sélectionner un segment de référence. Le choix de ce segment peut être guidé par des connaissances préalables de la pièce à analyser ou il peut simplement être effectué au hasard. Une analyse exhaustive peut également être effectuée. Une telle analyse équivaut à comparer tous les segments les uns avec les autres. La seconde étape de la démarche consiste à déterminer l'intersection entre les règles d'association du segment de référence et celles de chacun des autres segments. Le nombre de règles d'association communes donne alors le niveau de ressemblance des segments. Cette procédure est illustrée à la figure 6.19.



6.19. Processus d'analyse de la structure d'une pièce musicale

La durée des segments utilisés pour mener l'analyse influence grandement le procédé. Un segment de longue durée a tendance à devenir unique et par conséquent de n'être d'aucune utilité. En contrepartie, un segment de courte durée a tendance à contenir des événements sonores peu significatifs. L'utilisation de segments de courte durée peut faire en sorte qu'une multitude de motifs peu éloquents soit extraite. Plusieurs essais ont été effectués afin de pouvoir déterminer quelle est la durée à privilégier. Malheureusement, la durée optimale n'a pas été découverte. La portée de la durée semble variée en fonction de la nature de la musique analysée. Néanmoins, il a été possible d'obtenir des résultats intéressants sur une base régulière en fixant la durée à 5 secondes.

6.3.2. Ensemble de données

L'ensemble de données utilisées lors de cette expérimentation a été composé à partir de fichiers MP3 représentant des genres musicaux variés. Les pièces ont été sélectionnées de manière à ce que des genres différents à ceux contenus dans la collection GTZAN soient représentés afin qu'une autre perspective soit traitée. Au total, 1 000 pièces provenant de diverses collections privées ont été rassemblées pour créer cet ensemble de données.

6.3.3. Résultats et discussion

L'analyse d'une pièce musicale est une activité non triviale. La musique est un domaine qui possède plusieurs dimensions et plusieurs de ces dimensions ne sont pas formalisées. Les dimensions non formalisées sont caractérisées par l'absence de balise (Holland, 2000). Les problèmes liés au traitement de ces dimensions sont généralement mal définis ce qui laisse place à l'interprétation des résultats obtenus. Une solution envisagée doit pouvoir être révisée et modifiée au besoin (Cook, 1994). Il n'existe pas de bonne ou de mauvaise structure musicale. La composition ou l'expressivité dans la pratique musicale sont des aspects non formalisés de la musique. Pour ces deux aspects, il n'existe pas de problèmes concrets à résoudre. Tous au mieux, les problèmes liés à ces aspects peuvent se résumer à « composer quelque chose d'intéressant » (Levitt, 1985) ou « interpréter une musique avec émotion » (Holland, 2000). Cette considération ne signifie pas que ces aspects sont sans intérêt. Au contraire, ils peuvent servir à décrire ce qui différencie une musique d'une autre. Par conséquent, l'évaluation des résultats de cette expérimentation n'a pas été effectuée en égard avec une cible numérique à atteindre. L'évaluation a été portée sur la capacité de la démarche proposée à faire ressortir des propriétés

structurales intéressantes. La section suivante rapporte quelques observations capturées lors des analyses effectuées.

L'une des pièces analysées lors de cette expérimentation est la pièce Adrénaline²⁷ de la formation montréalaise Zébulon. La forme de l'onde sonore de cette pièce est donnée à la figure 6.20.



6.20. Forme de l'onde sonore de la pièce Adrénaline

Même s'il est difficile de l'observer en examinant uniquement la forme de l'onde sonore, une écoute de la pièce permet de la décomposer en 12 parties. La première partie de la pièce, d'une durée de 12 secondes, est une introduction jouée uniquement à la guitare. À la 13^e seconde de la pièce, la batterie, la basse et la voix sont ajoutées. Au même moment, la mélodie jouée à la guitare est altérée pour donner forme au premier couplet. À la 31^e seconde, la mélodie change pour créer une transition vers le refrain. Le refrain débute à la 43^e seconde et se termine à la 55^e seconde. La mélodie du refrain est la même que celle de l'introduction. Cependant, contrairement à l'introduction, les autres instruments accompagnent la guitare. La séquence couplet-transition-refrain est ensuite répétée. À 1 minute 30 secondes, une nouvelle mélodie est introduite pour

²⁷ <https://itunes.apple.com/ca/album/loeil-du-zig/id285451146?l=fr>

mener à un solo d'orgue. La mélodie qui accompagne le solo est la même que celle jouée lors de l'introduction et des refrains. Le solo débute à 1 minute 42 pour se terminer à 2 minutes 6 secondes. La même mélodie est jouée jusqu'à la fin de la pièce à l'exception d'une coupure à 2 minutes 26 secondes. Cette structure est illustrée à la figure 6.21.

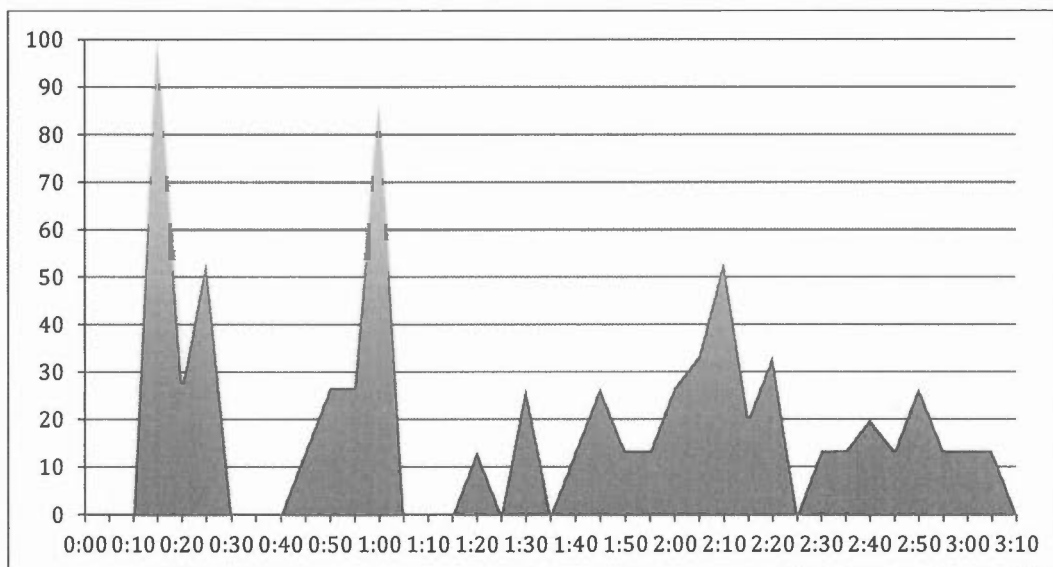


6.21. Découpage de l'onde sonore de la pièce Adrénaline

- 1) Introduction
- 2) Couplet #1
- 3) Transition #1
- 4) Refrain #1
- 5) Couplet #2
- 6) Transition # 2
- 7) Refrain #2
- 8) 2^e Transition
- 9) Solo d'orgue
- 10) Couplet altéré
- 11) Coupure
- 12) Final

L'analyse de la structure de la pièce a été effectuée en sélectionnant le segment débutant à la 15^e seconde comme segment de référence. L'extrait sonore contenu dans ce segment correspond à une portion du premier couplet. Les

règles d'association qui caractérisent ce segment ont été comparées à celles des autres segments. Le résultat de cette comparaison donné à la figure 6.22.

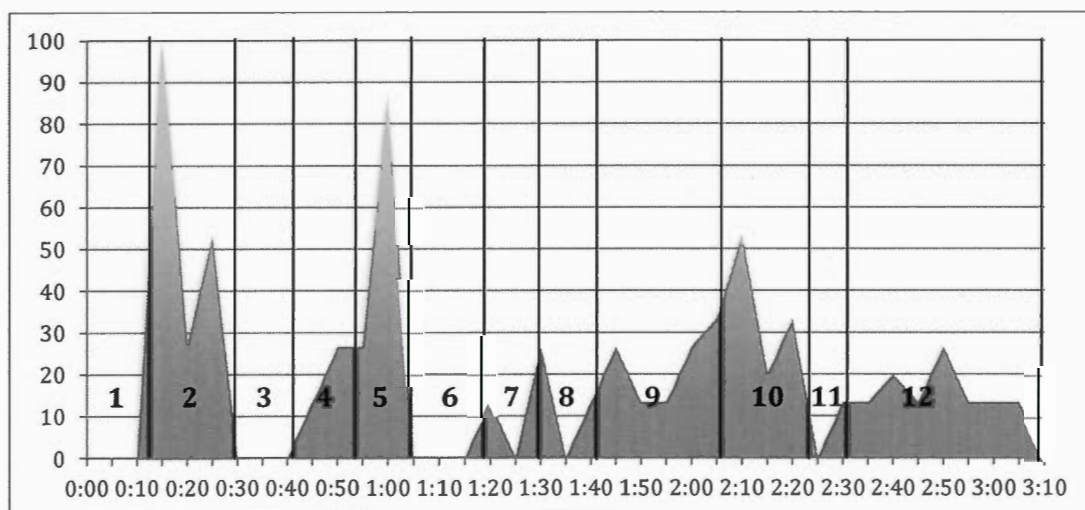


6.22. Graphique illustrant la structure de la pièce Adrenaline

L'abscisse représente le temps tandis que l'ordonnée indique le degré de similarité avec le segment de référence. Le segment de référence est situé à 0:15.

Un bref examen de la figure 6.22 permet de visualiser certaines structures de la pièce musicale. Les deux pointes les plus importantes soit celles situées à 0:15 et 1:00 représentent respectivement les couplets 1 et 2. Cette capacité à dégager précisément ces couplets est liée au choix du segment de référence. En effet, puisque le segment de référence contient un extrait du premier couplet, la correspondance entre ce segment et les autres segments qui réfère aux couplets est plus grande que pour tout autre segment. La pointe située à 2:10 représente également un couplet. Le degré de similarité est cependant moindre que pour la

pointe 1:00. Cet écart peut s'expliquer par le fait que la mélodie jouée à la guitare lors du 3^e couplet est différente de celle jouée lors des deux premiers couplets. Par conséquent, certaines règles d'associations dégagées des deux premiers couplets ne se retrouvent pas dans le 3^e couplet et vice versa. Dans ce contexte, il est normal que le seuil de similarité soit moindre. Une autre observation notable est que l'analyse effectuée a été en mesure de faire ressortir les transitions entre les couplets et les refrains. La transition #1 est représentée par le vide situé de 0:30 à 0:40 tandis que la transition #2 est représentée par le vide situé de 1:05 à 1:15. Ces observations sous-entendent que l'usage des règles d'association peut servir à dégager la structure de cette pièce musicale. La figure 6.23 illustre la structure de la pièce superposée au graphique généré.



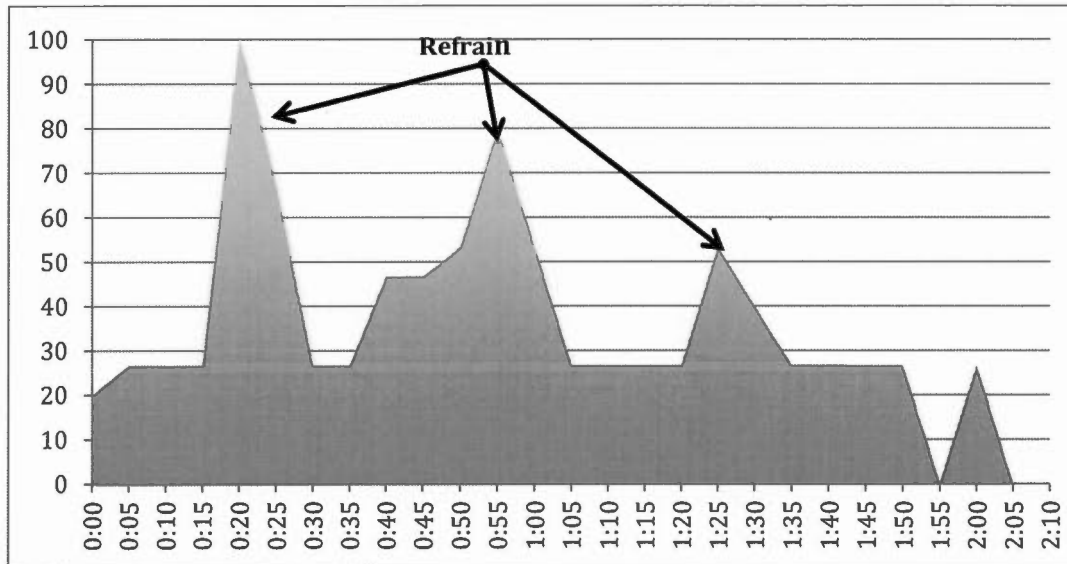
6.23. Découpage de la structure de la pièce Adrenaline

Certaines parties de la pièce musicale sont facilement repérables à l'intérieur du graphique généré. Par exemple, les 3 pointes les plus importantes correspondent aux couplets de la pièce.

Le repérage de refrain est une tâche liée à l'analyse de la structure des pièces musicales. Les essais effectués lors de l'expérimentation III ont permis de démontrer que les règles d'association peuvent servir à soutenir l'exécution de cette tâche. L'analyse de la pièce « Haircuts Uniform²⁸ » de la formation américaine Deleted Scenes constitue un exemple intéressant de cette application.

Le refrain est répété à trois occasions dans cette pièce. La première instance débute à 0:17 et se termine à 0:29. La seconde instance débute à 0:40 et se termine à 0:58. Finalement, la troisième instance débute à 1:20 et se termine à 1:32. L'analyse de la pièce a été effectuée en sélectionnant le segment 0:20 comme segment de référence. La figure 6.24 donne la représentation graphique de l'analyse réalisée.

²⁸ <https://itunes.apple.com/ca/album/lithium-burn/id853882223?l=fr>

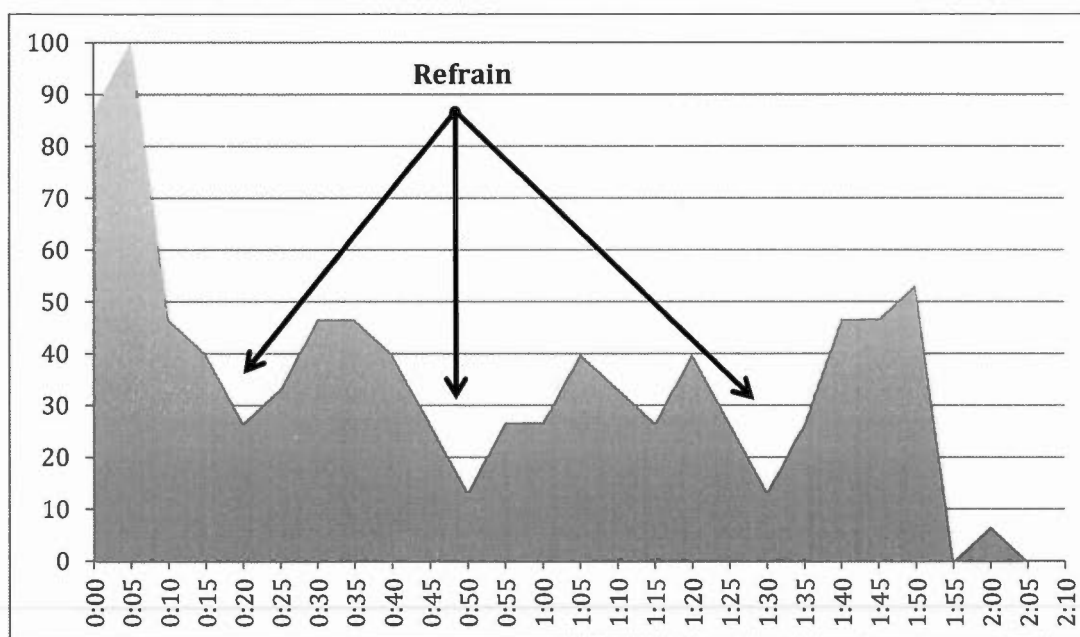


6.24. Graphique A de la pièce « Haircuts uniform »

L'analyse permet de localiser le refrain dans la pièce. Les trois pointes donnent la position du refrain. Ces pointes sont obtenues en utilisant le segment 0:20 comme segment de référence. Ce segment contient un extrait de la première instance du refrain.

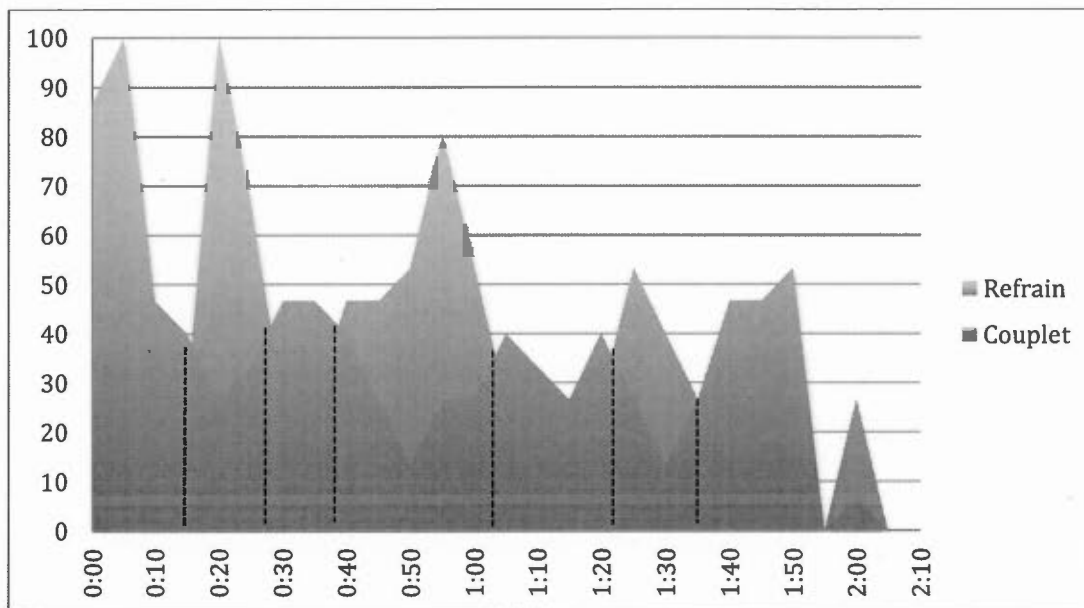
Les trois instances du refrain sont facilement repérables dans le graphique illustré à la figure 6.24. Ces instances sont représentées par trois importantes pointes sur le graphique. Cette analyse démontre que le refrain qu'une pièce musicale peut être dépeint à partir d'un ensemble de règles d'association et que la connaissance de ces règles peut être utilisé de manière à ce que le refrain puisse être localisé dans une pièce. Afin de contre valider les résultats obtenus, une analyse de la pièce a été effectuée en utilisant un segment de référence

contenant un extrait autre que le refrain. Les résultats de cette analyse sont illustrés à la figure 6.25. En examinant le graphique généré, on constate que les refrains sont représentés par des creux. Cette observation soutient l'hypothèse selon laquelle les instances du refrain sont caractérisées par des règles d'association qui ne sont pas contenues dans les couplets.



6.25. Graphique B de la pièce « Haircuts uniform »

L'analyse de la pièce a été effectuée en utilisant le segment 0:05 comme segment de référence. Les refrains sont représentés par des creux dans le graphique.

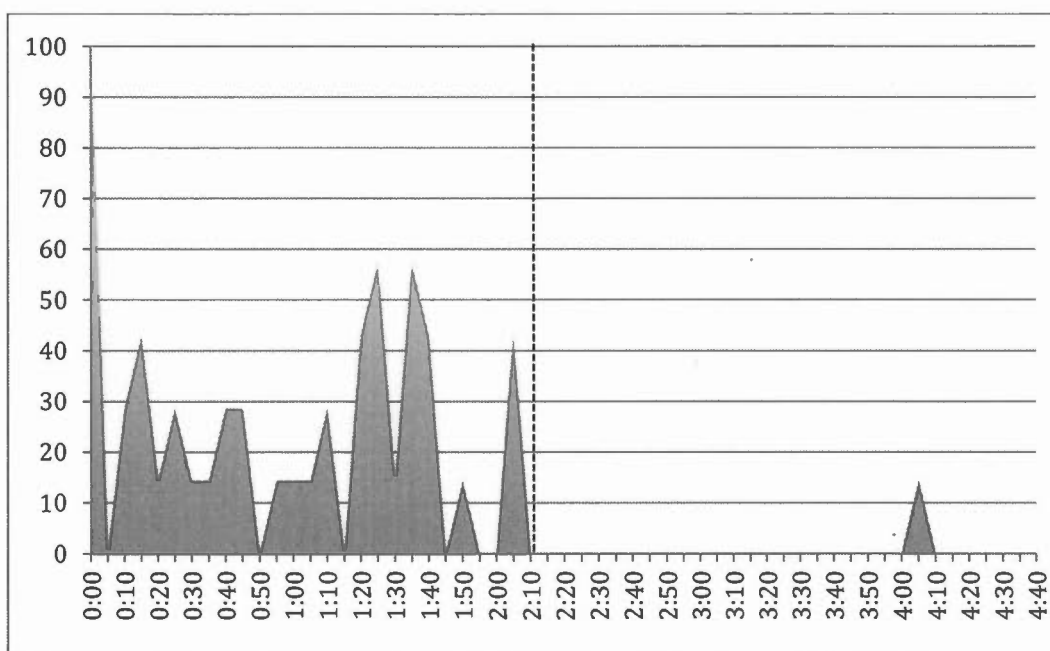


6.26. Superposition des graphiques A et B. de la pièce « Haircuts uniform »

Les zones où l'écart est le plus important correspondent quasi parfaitement aux positions du refrain dans la pièce.

La superposition des graphiques A et B illustrée à la figure 6.26 démontre qu'il existe une cohérence entre les deux analyses effectuées. En effet, les 3 pointes les plus importantes de la série qui exprime l'analyse où le segment de référence est un extrait du refrain et les 3 creux de la série qui exprime l'analyse où le segment de référence utilisé est un extrait du premier couplet concordent. Les zones où l'écart entre les deux séries est le plus important correspondent quasi parfaitement aux positions du refrain dans la pièce. Cette capacité à localiser les refrains peut être transposée de sorte qu'il soit possible de localiser différentes séquences dans une pièce musicale.

Une autre pièce analysée lors de cette expérimentation est la pièce « My body is a cage » de la formation Arcade Fire²⁹. Cette pièce débute par la phrase « My body is a cage » chantée a cappella. Dans la première partie de la pièce, cette phrase est répétée à 5 reprises. À 2 minutes 10 secondes, l'atmosphère de la pièce devient plus lourde. Le rythme change partiellement et la mélodie jouée à l'orgue devient plus instance. La figure 6.27 donne le graphique obtenu suite à une analyse de la pièce effectuée en utilisant le segment 0:00 comme segment de référence.



6.27. Graphique de la pièce « My body is a cage »

L'analyse de la pièce a été effectuée en utilisant le segment 0:00 comme segment de référence.

²⁹ <https://itunes.apple.com/ca/album/neon-bible/id216167680?l=fr>

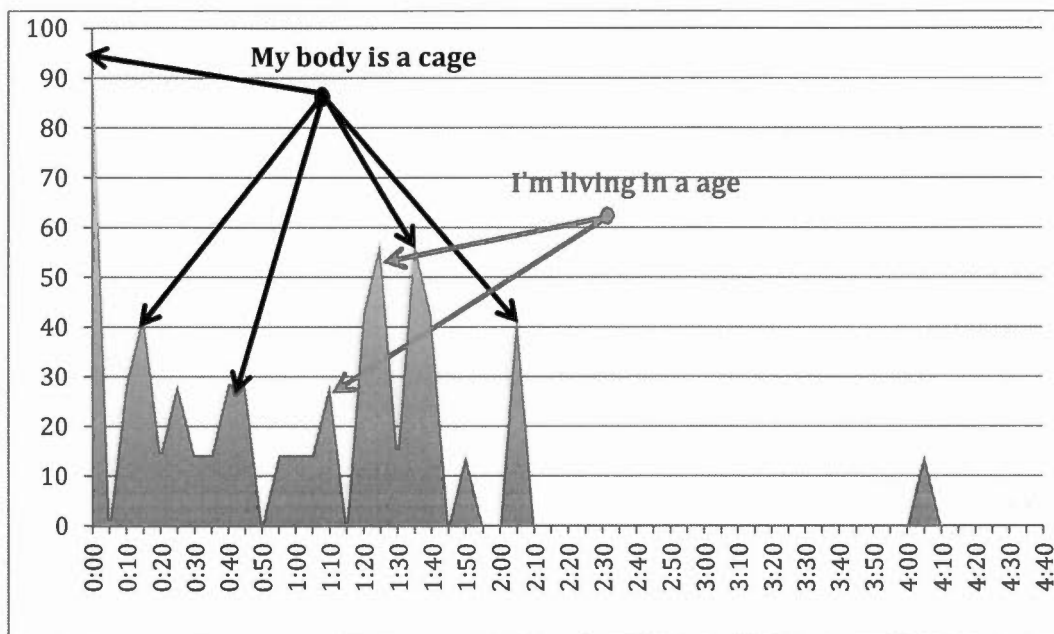
La première observation qu'il est possible de noter suite à l'examen de la figure 6.27 est que le changement d'atmosphère perceptible à partir de la deuxième moitié de la pièce est visible sur le graphique. Ce changement est observé à 2:10. De manière générale, la première partie de la pièce est plus épurée que la seconde. La voix y est mise à l'avant-plan. Dans la seconde partie de la pièce, les instruments occupent plus une place plus importante ce qui ajoute de la texture à la pièce.

Il semble avoir un lien direct entre cette particularité et le nombre de règles extraites. En moyenne, 26 règles d'association sont extraites des segments de la première partie de la pièce tandis que cette moyenne est 52 pour la seconde partie. Le tableau 6.11 donne le nombre de règles d'association extraites pour chacun des segments.

1 ^{re} partie		2 ^e partie	
Segment	Nombre de règles d'association	Segment	Nombre de règles d'association
0:00	6	2:15	67
0:05	10	2:20	30
0:10	14	2:25	126
0:15	4	2:30	74
0:20	12	2:35	24
0:25	12	2:40	46
0:30	4	2:45	26
0:35	6	2:50	26
0:40	10	2:55	40
0:45	10	3:00	42
0:50	4	3:05	60
0:55	28	3:10	84
1:00	56	3:15	90
1:05	124	3:20	52
1:10	18	3:25	58
1:15	22	3:30	44
1:20	30	3:35	30
1:25	26	3:40	30
1:30	20	3:45	32
1:35	34	3:50	28
1:40	16	3:55	42
1:45	14	4:00	24
1:50	36	4:05	42
1:55	50	4:10	50
2:00	56	4:15	34
2:05	38	4:20	38
2:10	46	4:25	26
		4:30	80
		4:35	70
Moyenne :	26	Moyenne :	52

6.11. Nombre de règles d'association extraites par segment de 5 secondes

Une deuxième observation considérable est que les 5 occasions dans la première moitié de la pièce où la phrase « My body is a cage » est chantée sont toutes représentées par une pointe sur le graphique. Ces pointes sont situées à 0:00, 0:15, 0:45, 1:40 et 2:08. La pointe située à 1:25 illustre un fort degré de similarité entre le segment situé à cette période et le segment de référence. Pourtant, l'écoute du segment à 1:25 permet de constater l'absence de la phrase titre. Cependant, même si les paroles sont différentes il existe réellement une similarité avec le segment de référence. En effet, la phrase « I'm living in a age » y est chantée avec le même aire que celle contenue dans le segment de référence. De plus, le dernier mot de chacune des phrases (« cage » et « age ») diffère seulement d'un phonème. En fait, toutes les pointes au-delà de 25%, à l'exception de celle située à 0:25, représentent des extraits sonores dont l'air est la même que celle contenue dans le segment de référence. Ces résultats suggèrent que le système proposé peut être calibré de manière à ce qu'il puisse faire ressortir des répétitions au niveau du chant. Par contre, ces répétitions peuvent être détectées uniquement lorsque la voix est mise à l'avant-plan.



6.28. Localisation de la phrase « My body is a cage » et « I'm living in a cage »

Les résultats présentés précédemment sous-entendent que les règles d'association peuvent être utilisées pour dégager la structure de pièces musicales. Cette capacité peut servir à localiser les refrains dans une pièce musicale et même être employée pour détecter des motifs récurrents dans le chant. La principale difficulté liée à l'usage des règles d'association dans ce contexte réside dans le choix de la valeur des paramètres. En effet, la solution proposée est sensible à la variation de la valeur attribuée à ses différents paramètres. Par exemple, le choix du segment de référence influence substantiellement les structures dégagées. Cependant, lorsque calibrée adéquatement, l'approche proposée est en mesure de dégager des motifs qui sont en mesure de représenter la structure d'une pièce musicale.

Les résultats de l'expérimentation III combinés à ceux de l'expérimentation II démontrent la portabilité de règles d'association. Avec peu d'effort il est possible d'adapter les règles d'association afin de mettre en œuvre différentes stratégies pouvant servir l'exploration de données musicales.

6.4. Expérimentation IV : recommandation d'artistes

Les expérimentations II et III ont été consacrées au traitement de données représentant le signal sonore. Dans le cadre de l'expérimentation IV, des données textuelles périphériques aux pièces musicales ont été utilisées pour former les transactions. Lors d'une recherche de musique, les données textuelles périphériques aux documents audio sont également consultées. Par conséquent, un document peut être écarté ou avantagé suite à l'évaluation des données périphériques qui l'accompagnent. Un outil d'aide à l'exploration de données musicales flexible doit donc être en mesure de considérer à la fois les données brutes et les données textuelles périphériques aux documents audio. L'expérimentation IV vise à valider que les règles d'associations peuvent être appliquées à des descripteurs autres que des descripteurs de bas niveau.

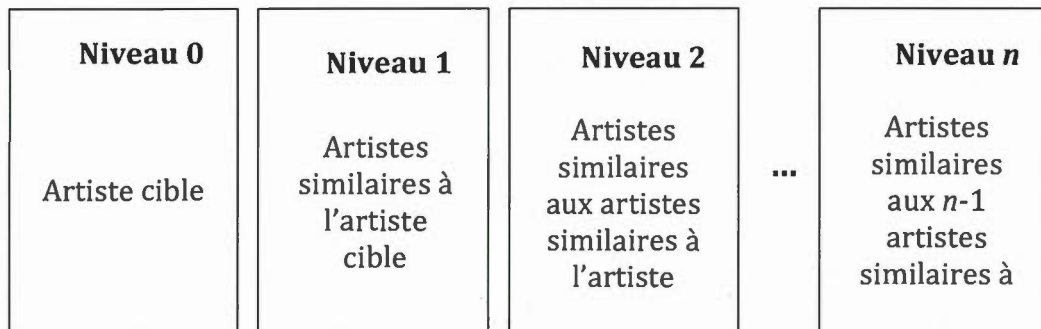
La majorité du temps, le nom des interprètes accompagne les enregistrements sonores. Il est fréquent qu'une liste d'artistes similaires soit également présentée. La navigation au sein des multiples documents audio contenus dans une base de données peut être influencée par cette liste. Les systèmes de recommandation de musique exploitent ce principe. Comme il a été présenté au chapitre 4, Swathi et Reddy ont proposé d'utiliser les règles d'associations afin d'améliorer les systèmes de recommandation de musique (Swathi et Reddy,

2014). Leurs travaux ont porté plus spécifiquement sur le développement d'une méthode destinée à améliorer le filtrage collaboratif. Les scores attribués aux différentes pièces musicales par les utilisateurs sont employés comme item pour créer les transactions. Le problème abordé consiste à prédire quelles pièces peuvent intéresser un utilisateur en fonction des scores que cet utilisateur a attribué à d'autres pièces contenues dans la base de données. L'approche proposée par Swathi et Reddy est intéressante, car elle explore l'usage des règles d'associations pour traiter des données périphériques aux documents audio. Cependant, les auteurs donnent peu de détails sur les résultats qu'ils ont obtenus lors de leurs expérimentations. L'expérimentation IV consiste donc à approfondir l'exploration de l'utilisation des règles d'association pour le traitement des données périphériques.

6.4.1. Méthodologie

La méthode développée lors de cette expérimentation a comme objectif de bonifier automatiquement la liste des artistes similaires à un artiste donné. Contrairement à la méthode proposée par Swathi et Reddy, elle ne vise pas à soutenir le filtrage collaboratif. Le nombre potentiellement élevé d'utilisateurs impliqués dans un système de recommandation basé sur le filtrage collaboratif combiné au nombre important de pièces accessibles dans les bases de données audio peut mener à une explosion combinatoire. L'effort computationnel nécessaire au traitement de ces informations est tel que ce traitement devient à toute fin impraticable. Afin de pallier à cet obstacle, il est nécessaire de limiter l'espace de recherche.

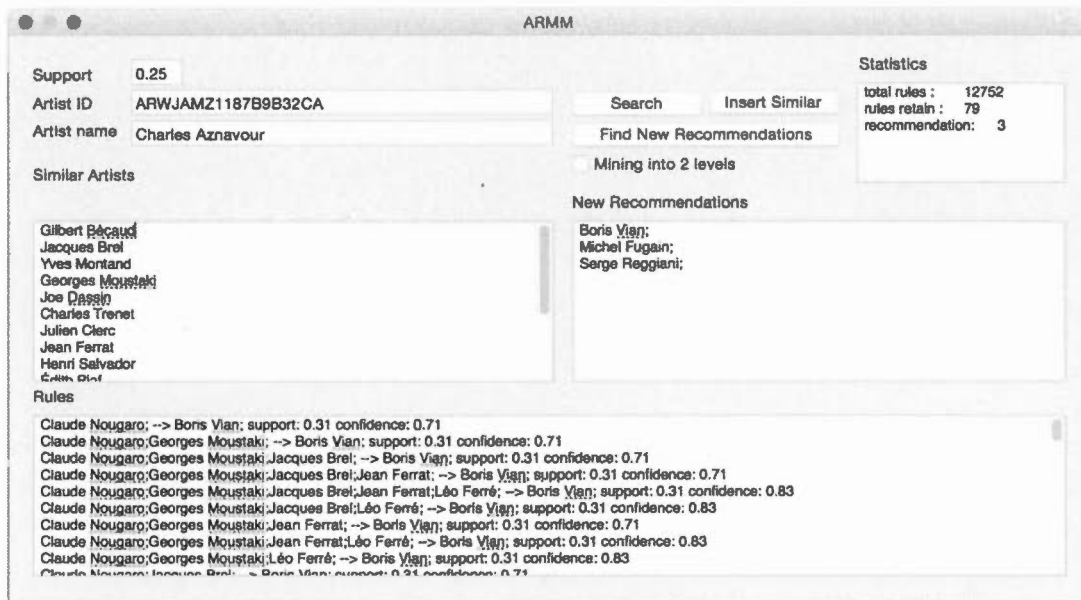
La méthode proposée consiste à utiliser la liste des artistes similaires à un artiste ciblé pour former les transactions. Pour ce faire, un artiste est préalablement sélectionné. La liste des artistes qui lui sont similaires est ensuite obtenue à partir d'un appel à un service web (les détails sont donnés à la section 5.3.2). L'artiste ciblé est ajouté à cette liste pour former une première transaction. Pour chacun des artistes contenus dans la transaction, le procédé est répété de manière à créer un ensemble de transactions. Généralement, le nombre d'artistes similaires retournés est raisonnable. En moyenne, ce nombre se situe entre 10 et 15. Le contrôle de l'espace de recherche se fait en limitant le nombre de niveaux forés. L'artiste cible est considéré comme appartenant au niveau 0 tandis que les artistes qui lui sont similaires sont considérés comme appartenant au niveau 1. Cette structure est illustrée à la figure 6.29.



6.29. Profondeur des transactions

Limiter le nombre de niveaux considéré permet d'exercer un contrôle sur l'espace de recherche. Ce contrôle réduit le risque d'explosion combinatoire.

Les essais effectués lors de la conception du prototype ont permis de découvrir que le forage peut être limité aux deux premiers niveaux. Au-delà, la plupart du temps, aucune nouvelle recommandation n'est faite. La raison derrière ce phénomène est que plus il y a de niveau de considéré, plus il y a d'itemsets possibles. Par conséquent, le nombre d'itemsets fréquents tend à diminuer. Un autre facteur qui pousse à limiter le nombre de niveaux forés est que le temps nécessaire à l'extraction des règles croît en fonction du nombre de niveaux ajoutés. La figure 6.30 illustre une capture d'écran du prototype développé.



6.30. Capture d'écran du module d'exploration d'artistes similaires

Le prototype limite le forage à deux niveaux. Les essais effectués lors de développement de ce prototype ont permis de découvrir que la plupart du temps aucune nouvelle recommandation n'est faite au-delà de 2 niveaux.

6.4.2. Ensemble de données

Les services web fournis par Echonest ont été utilisés à titre de source de données. Echonest est une entreprise dédiée aux traitements intelligents de la musique. Elle a été fondée par Tristan Jehan et Brian Whitman deux chercheurs intéressés par le traitement automatisé de la musique (Jehan, 2005 ; Whitman, 2005). Les services offerts par cette entreprise sont exploités aussi bien par le milieu des affaires que par des chercheurs intéressés au traitement de la musique (Schindler et Rauber, 2012 ; Bertin-Mahieux et al., 2011). Des entreprises telles que Cola Cola, Intel, Microsoft et Reebok utilisent les services d'Echonest pour rehausser l'expérience qu'elles offrent à leurs clients. En avril 2015, Echonest disposait d'information sur plus de 3 millions d'artistes et 36 millions d'enregistrements sonores ! Echonest permet aux chercheurs d'accéder gratuitement à ses données. L'accès aux données se fait à travers l'utilisation de services web. L'entreprise a développé plusieurs services web spécialisés dans la recherche d'informations précises. C'est en interrogeant l'un de ces services web que la liste des artistes jugés similaires à un artiste cible peut être obtenue. L'exemple 5.1 donne l'URL permettant d'obtenir la liste des artistes similaires au chanteur français Charles Aznavour.

`http://developer.echonest.com/api/v4/artist/similar?api_key=[API_KEY]&id=ARWJAMZ1187B9B32CA`

Exemple 5.1 : Exemple d'un appel au service web d'Echonest permettant d'obtenir la liste des artistes jugés similaire à Charles Aznavour où [API_KEY] correspond à un jeton doit être remplacé par une clé d'enregistrement unique fournie gratuitement par Echonest et la chaîne de caractères ARWJAMZ1187B9B32CA représente un l'identifiant unique qui réfère à Charles Aznavour

6.4.3. Résultats et discussion

La première évaluation effectuée a porté sur la recherche d'artistes similaires au chanteur français Jacques Brel. Au total, 15 artistes jugés similaires ont été retournés par le service web d'Echonest. Ces artistes sont listés dans la table 6.12.

Artiste similaire
Yves Montand
Charles Aznavour
Gilbert Bécaud
Georges Moustaki
Serge Gainsbourg
Charles Trenet
Édith Piaf
Georges Brassens
Jean Ferrat
Julien Clerc
Boris Vian
Claude Nougaro
Joe Dassin
Léo Ferré
Henri Salvador

6.12. Artistes similaires à Jacques Brel selon Echonest

L'écoute de quelques pièces musicales interprétées par les artistes listés dans le tableau 6.12 permet de confirmer qu'il existe bel et bien un lien de similarité entre ces artistes et l'artiste cible. Pour chacun de ces artistes, la liste des artistes qui leur est similaire a été recherchée. Au total, 76 artistes distincts ont été retournés. Lorsque le niveau 2 a été considéré, ce nombre est passé à 147. Dans

ce contexte, le problème à résoudre consiste à dégager les artistes qui sont les plus susceptibles de ressembler à l'artiste cible. Plutôt que de sélectionner aléatoirement des artistes, la sélection a été orientée par des règles d'association. L'hypothèse soutenue est que plus des artistes apparaissent fréquemment ensemble, plus la probabilité qu'ils soient similaires est grande.

Une première évaluation a été effectuée en fixant les mesures de support et de confiance à 25%. Lors de cette évaluation, seuls les artistes des niveaux 1 et 2 ont été considérés. Au total, 51 628 règles d'associations ont été produites. Tous les artistes contenus dans le conséquent des règles qui ne figuraient pas la liste des artistes du niveau 1 ont été retenus comme candidat. Ainsi, 8 artistes ont été dégagés. Ces artistes ont été répertoriés dans 270 des 51 628 règles d'association produites. Le tableau 6.13 donne la liste des artistes jugés similaires à Jacques Brel en fonction des règles d'association.

Artiste similaire
Dalida
Juliette Gréco
Maxime Le Forestier
Michel Fugain
Michel Sardou
Mouloudji
Serge Reggiani
Tino Rossi

6.13. Artistes jugés similaires en fonction des règles d'association.

Il est difficile de déterminer si les artistes listés dans le tableau 6.13 forment une solution optimale. Toutefois, l'écoute de quelques pièces interprétées par ces

artistes permet de conclure qu'ils constituent une liste cohérente avec celle du niveau 1. Les artistes dégagés à l'aide des règles d'association peuvent donc être utilisés pour bonifier la liste d'artistes similaires. Le tableau 6.14 donne la liste des règles d'association utilisées pour dégager les artistes similaires.

Antécédent	Conséquent	SUP	CONF
Charles Aznavour, Charles Trenet, Gilbert Bécaud, Jacques Brel, Yves Montand	Juliette Gréco	0,27	0,57
Charles Aznavour, Charles Trenet, Gilbert Bécaud, Jacques Brel, Yves Montand	Tino Rossi	0,27	0,57
Charles Aznavour, Georges Moustaki, Gilbert Bécaud, Jacques Brel, Joe Dassin, Yves Montand	Dalida	0,27	0,8
Charles Aznavour, Georges Moustaki, Gilbert Bécaud, Jacques Brel, Jean Ferrat	Serge Reggiani	0,33	0,62
Charles Aznavour, Georges Moustaki, Gilbert Bécaud, Jacques Brel, Julien Clerc	Michel Fugain	0,27	0,67
Charles Aznavour, Georges Moustaki, Gilbert Bécaud, Jacques Brel, Yves Montand	Michel Sardou	0,27	0,57
Charles Aznavour, Gilbert Bécaud, Henri Salvador, Jacques Brel	Mouloudji	0,27	0,57
Georges Moustaki, Jacques Brel, Jean Ferrat	Maxime Le Forestier	0,27	0,44

6.14. Règles d'association utilisées pour dégager de nouveaux artistes

Les essais effectués ont permis de démontrer que le nombre d'artistes dégagés peut être contrôlé en fixant différentes valeurs de support. Le tableau 6.15 donne la relation entre la mesure de support et le nombre d'artistes dégagés.

Support	Nombre de règles produites	Nombre de règles informatives	Nombre d'artistes dégagés
0.50	222	0	0
0.40	1922	9	1
0.35	11762	43	3
0.30	11762	43	3
0.25	51628	270	8
0.20	343966	1247	11
0.15	343966	1247	11
0.10	7170866	8418	22

6.15. Relation entre support et nombre d'artistes dégagés

Les artistes sont dégagés à partir des règles informatives. Une règle est considérée informative lorsque son conséquent contient un artiste qui n'est pas contenu dans le niveau 1.

Lors de tous les essais effectués, l'écoute de quelques pièces des artistes dégagés a permis de confirmer qu'ils étaient similaires à artistes ciblés. Par exemple, Bill Haley & His Comets, Buddy Holly & The Crickets, Buddy Knox et Eddie Cochran ont été dégagés comme étant des artistes similaires à Elvis Presley, Neil Young et Steve Miller Band ont été dégagés comme étant des artistes similaires à Tom Petty et Antonio Vivaldi, Franz Liszt, Georg Friedrich Händel, Gustav Mahler, Hector Berlioz, Jean Sibelius et Robert Schumann ont été dégagés comme étant des artistes similaires à Ludwig van Beethoven. Dans tous les cas, ces artistes n'étaient pas présentés initialement comme des artistes similaires à l'artiste cible. Le fait de limiter l'espace de recherche fait en sorte que la méthode proposée est en mesure de découvrir dans la majorité des cas de nouvelles relations dans un temps raisonnable. La qualité des résultats produits dépend majoritairement de la pertinence des données forcées. Bien que les

évaluations aient été effectuées uniquement en considérant des listes d'artistes jugés similaires, il est facile d'extrapoler comment les règles d'association peuvent être appliquées à d'autres descripteurs textuels comme le genre, la région, l'année, etc.

Les résultats de cette expérimentation jumelés à ceux des expérimentations II et III démontrent que les règles d'association peuvent opérer sur des descripteurs appartenant à différents niveaux d'abstraction pour mettre en œuvre différentes stratégies liées à l'exploration des données musicales.

6.5. Discussion

L'objectif général de ce projet de thèse étant de démontrer la flexibilité et la portabilité des règles d'association dans un contexte d'aide à l'exploration de données musicales, chacune des expérimentations a été effectuée en lien avec cet objectif. Les expérimentations réalisées ont mené à des résultats intéressants permettant de soutenir les hypothèses qui découlent de l'objectif général.

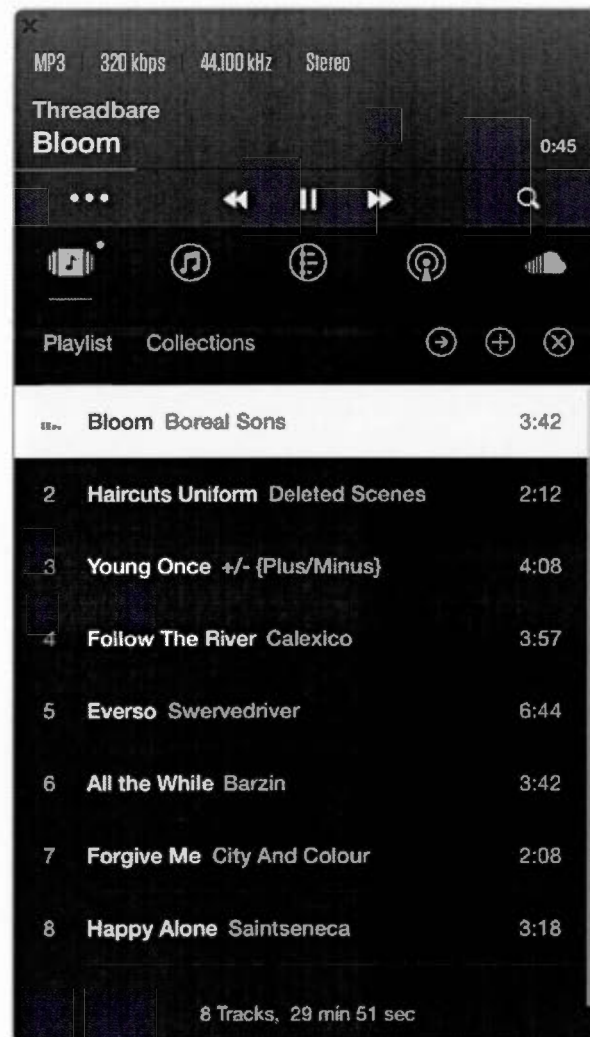
La première hypothèse qui a été posée est que l'expérience que procure l'écoute d'une musique dépend fortement du traitement de plusieurs types d'informations dont le niveau d'abstraction varie. Il est présumé que ce sont ces traitements qui permettent entre autres de mesurer la similarité entre deux pièces musicales et éventuellement de catégoriser la musique. En considérant cette hypothèse, il est probable que les outils actuels puissent être limités par le fait qu'ils opèrent seulement sur des descripteurs appartenant à un seul niveau d'abstraction. Le descripteur le plus utilisé pour représenter la musique est le genre musical. L'expérimentation I a été réalisée afin de mesurer les limites liées

à l'usage exclusif de ce descripteur. Les résultats obtenus lors de cette expérimentation ont souligné le caractère subjectif du genre musical. L'ajout de descripteur doit être perçu comme un moyen d'enrichir la manière dont la musique est décrite.

La seconde hypothèse soutenue dans cette thèse est que les relations entre les descripteurs utilisés pour décrire une musique sont tout aussi importantes que la présence ou l'absence de ces descripteurs. Les résultats des expérimentations II et III ont confirmé l'utilité des relations de cooccurrence dégagées par les règles d'association dans un contexte d'exploration de données musicales. Il a été démontré que ces relations peuvent être utilisées afin d'évaluer le degré de similarité entre des pièces musicales et même servir à dégager leur structure.

Finalement, la troisième hypothèse qui a été vérifiée lors des expérimentations est que l'utilisation des règles d'association peut, d'une part, faciliter la mise en œuvre de diverses stratégies utiles à l'exploration de données musicales et, d'autre part, servir à traiter l'information provenant de différents niveaux d'abstraction. Cette hypothèse a été validée à travers les expérimentations II, III et IV. Dans le cadre de ces expérimentations, des mécanismes d'aide à la classification, à l'analyse de la structure des pièces musicales et à la recherche d'artistes similaires ont été développés. Le développement de ces mécanismes témoigne de la portabilité des règles d'association. Il témoigne également de la flexibilité des règles d'association. Les mécanismes développés ont été appliqués à des descripteurs appartenant à différents niveaux d'abstraction. Les expérimentations II et III ont démontré que les règles d'association peuvent opérer sur des descripteurs de bas niveau tandis que l'expérimentation IV a démontré qu'elles peuvent opérer sur des descripteurs de niveau intermédiaire. Même si les expérimentations effectuées n'ont pas porté spécifiquement sur

l'usage des descripteurs de haut niveau, elles donnent tous les fondements permettant leur intégration. Les descripteurs de haut niveau réfèrent aux préférences et aux motivations des utilisateurs. Comme il a été présenté dans le chapitre 4, il est possible d'utiliser les règles d'association pour extraire des connaissances cachées dans un ensemble de données dans la mesure où des transactions peuvent être définies dans cet ensemble. C'est principalement en misant sur cet aspect que les informations de haut niveau peuvent être intégrées. Par exemple, les listes de lectures créées par un utilisateur peuvent servir à préciser les préférences de cet utilisateur en plus de renseigner sur la manière dont il agence ses préférences. Dans ce contexte, les pièces qui composent les listes de lectures peuvent servir à définir les transactions (6.31). Ainsi, les mêmes méthodes que celles utilisées lors des expérimentations présentées peuvent être employées pour extraire des connaissances cachées dans les préférences des utilisateurs. Ces connaissances peuvent ensuite être utilisées pour personnaliser l'exploration de données musicales. Elles peuvent servir entre autres à bonifier les listes de lecture des utilisateurs, à organiser leurs bibliothèques musicales, à raffiner les résultats de leurs recherches, etc. Ceci représente seulement un échantillon des applications possibles.



6.31. Lecteur audio

La méthodologie proposée peut être appliquée aux listes de lecture créées par un utilisateur afin de dégager des connaissances cachées dans ses préférences. Dans ce contexte, la liste de lecture est utilisée comme ensemble de données et les pièces qui composent cette liste sont utilisées pour définir les transactions.

CHAPITRE VII

CONCLUSION

Cette thèse s'est articulée autour du problème de l'exploration des données musicale. L'engouement pour les bases de données composées de documents audio et les problématiques qui en découlent ont motivé ce choix. Certaines bases de données peuvent être composées de plusieurs millions de documents audio. Lorsqu'un utilisateur a une idée précise de ce qu'il souhaite accomplir, il peut parvenir à s'orienter parmi toutes ces données. Autrement, l'exploration des données musicales peut devenir une tâche complexe qui demande un effort considérable. L'exploration de données musicales ne se limite pas à l'exploration d'un ensemble de documents audio. Elle peut être dirigée à l'intérieur même des documents. Plusieurs stratégies ont été proposées afin de soutenir l'exploration des données musicales. Cependant, chacune des stratégies proposées a tendance à couvrir seulement une facette de l'exploration des données musicales. Elles proposent des mécanismes spécifiques que ce soit pour organiser, analyser ou accéder aux données. Par conséquent, l'exploration de données musicales nécessite l'utilisation de plusieurs outils complémentaires.

La musique peut être abordée selon différentes perspectives. Plusieurs évidences suggèrent que l'exploration de données musicales peut être bonifiée lorsque des descripteurs appartenant à divers niveaux d'abstraction sont considérés. Le fait de considérer des descripteurs variés mène ultimement à la construction d'une définition riche et significative. Une telle définition peut alors

être utilisée pour explorer les différentes facettes de la musique. À travers les chapitres de cette thèse, il a été démontré que les descripteurs utilisés pour décrire la musique peuvent être groupés de manière à former 3 niveaux d'abstraction. Le premier niveau contient les descripteurs de bas niveau. Ces descripteurs sont utilisés pour décrire l'onde sonore qui véhicule la musique. Le deuxième niveau réfère aux étiquettes apposées à des séquences particulières de sons. Ces descripteurs renseignent sur les conventions sociales et culturelles qui dictent la manière dont la musique est décrite. La plupart du temps, ce sont des étiquettes textuelles. Finalement, les descripteurs du troisième niveau correspondent aux indices qui permettent de cibler les préférences et les motivations des utilisateurs. C'est uniquement en considérant les descripteurs de ce niveau que la valeur réelle accordée par un individu à une musique peut être capturée.

Les travaux de recherche présentés dans ce document ont porté sur l'usage des règles d'association pour soutenir l'exploration de données musicales. Les règles d'association traduisent des relations de cooccurrence entre des données. Elles ont la capacité de faire ressortir des connaissances cachées à l'intérieur de vastes ensembles de données. Le travail effectué a été de démontrer la flexibilité et la portabilité des règles d'association. La démonstration que les règles d'association possèdent ces qualités a conduit à la conceptualisation d'une architecture modulaire facilitant la mise en œuvre de diverses stratégies pour soutenir l'exploration de données musicales. Les expérimentations effectuées ont permis de confirmer que les règles d'association sont flexibles et portables. Lors de ces expérimentations, elles ont été utilisées pour supporter la classification de documents audio, soutenir l'analyse de pièces musicales et pour bonifier des listes d'artistes jugés similaires. En plus d'avoir été utilisées pour implémenter différentes stratégies liées à l'exploration de données musicales,

les règles d'association ont été utilisées pour manipuler des descripteurs appartenant à différents niveaux d'abstraction. En somme, il a été démontré que les règles d'association peuvent aussi bien manipuler des descripteurs représentant l'onde sonore que des descripteurs textuels utilisés pour décrire la musique.

7.1. Perspectives

Cette recherche s'ajoute aux efforts déployés pour améliorer le traitement des données musicales. Elle propose l'usage d'une mécanique unique pour mettre en œuvre plusieurs stratégies vouées à l'exploration de données musicales. Cette mécanique repose sur la flexibilité et la portabilité des règles d'association. La valeur ajoutée de la solution proposée est que les règles d'association ne sont pas opaques. Par conséquent, il est facile de les interpréter et de les adapter. En plus de promouvoir le développement des stratégies exposées dans ce document, les perspectives du travail effectué sont les suivantes :

- **Améliorer la classification automatisée de données musicales.**
Sélectionner les bons descripteurs représente l'une des principales difficultés associées au développement des systèmes de classification automatisée de données musicales. L'opacité des méthodes actuelles rend difficile l'évaluation de la contribution des descripteurs dans le processus de classification. La méthode proposée peut servir à mesurer la contribution des descripteurs. Par conséquent, elle peut être utilisée pour faciliter l'identification des descripteurs qui sont les plus discriminants.

- **Générer des index pour les documents audio**

L'antécédent et le conséquent des règles d'association peuvent être utilisés pour indexer les documents sonores. Le potentiel de cette démarche est illustré en annexe (voir annexe A)

- **Dégager des connaissances entre des descripteurs de différents niveaux.** Les expérimentations réalisées dans le cadre de cette thèse ont démontré que les règles d'association peuvent être utilisées pour dégager des relations entre des descripteurs de bas niveau et des genres musicaux. Cette capacité à établir des relations entre des descripteurs de différents niveaux d'abstraction demande à être exploitée afin de découvrir si des descripteurs de bas niveau peuvent être associés aux préférences d'un utilisateur par exemple.

- **Soutenir l'identification des différentes versions d'une pièce musicale.** Les règles d'association sont en mesure de faire ressortir la structure des pièces musicales. Il est possible que cette capacité puisse servir à soutenir l'identification des différentes versions d'une pièce musicale.

ANNEXE A

EXEMPLE DE REQUÊTE SQL EXPLOITANT LES RÈGLES D'ASSOCIATION

Cet exemple illustre l'utilité des connaissances extraites à l'aide des règles d'association comme critères pour effectuer des recherches dans une base de données. Les règles d'association du tableau 6.8 (6.8. Règles d'association de deux pièces présumées CLASSIQUE) sont utilisées dans cet exemple.

Règles 1 : '797;' → '517;'

Règles 2 : '323;' → '517;'

Règles 3 : '625;' → '517;'

Règles 4 : '410;' → '517;'

L'antécédent et le conséquent de quelques règles communes aux deux pièces musicales représentées dans le tableau 6.8 sont utilisés comme critères de recherche. Plusieurs règles peuvent être exploitées en imbriquant les requêtes SQL.

Cet exemple démontre que le pourcentage des documents classiques retournés augment proportionnellement au nombre de règles discriminantes considérées.

L'ensemble de données GTZAN a été utilisé pour alimenter la base de données utilisée dans cet exemple. 10 genres musicaux sont représentés uniformément par 100 documents audio. Les genres représentés sont le blues, le classique, le

country, le disco, le hip-hop, le jazz, le métal, la musique populaire, le reggae et le rock.

EXEMPLE DE REQUÊTE EXPLOITANT 1 SEULE RÈGLE D'ASSOCIATION

```
SELECT DISTINCT
  AUDIO_FILE
FROM
  ASSOCIATION_RULES
WHERE
  ANTECEDANT = '797;' AND CONSEQUENT = '517'
```

RÉSULTATS POUR LA REQUÊTE EXPLOITANT 1 SEULE RÈGLE D'ASSOCIATION

Règles 1 : '797;' → '517;'

Genre	Document	Pourcentage
BLUES	20	17%
CLASSIQUE	72	63%
COUNTRY	8	7%
DISCO	6	5%
HIP-HOP	1	1%
JAZZ	6	5%
MÉTAL	1	1%
POP	1	1%
REGGAE	0	0%
ROCK	0	0%
Total	115	

EXEMPLE D'UNE REQUÊTE EXPLOITANT 2 RÈGLES D'ASSOCIATION

```

SELECT DISTINCT
  AUDIO_FILE
FROM
  ASSOCIATION_RULES
WHERE
  ANTECEDANT = '323;' AND CONSEQUENT = '517;' AND
AUDIO_FILE IN
  (SELECT DISTINCT
    AUDIO_FILE
  FROM
    ASSOCIATION_RULES
  WHERE
    ANTECEDANT = '797;' AND CONSEQUENT = '517;')

```

RÉSULTATS POUR LA REQUÊTE EXPLOITANT 2 RÈGLES D'ASSOCIATION

Règles 1 : '797;' → '517;'

Règles 2 : '323;' → '517;'

Genre	Document	Pourcentage
BLUES	14	16%
CLASSIQUE	60	68%
COUNTRY	5	6%
DISCO	3	3%
HIP-HOP	1	1%
JAZZ	4	5%
MÉTAL	1	1%
POP	0	0%
REGGAE	0	0%
ROCK	0	0%

Total

88

RÉSULTATS DE LA REQUÊTE EXPLOITANT 3 RÈGLES D'ASSOCIATION

Règles 1 : '797;' → '517;'

Règles 2 : '323;' → '517;'

Règles 3 : '625;' → '517;'

Genre	Document	Pourcentage
BLUES	13	16%
CLASSIQUE	56	69%
COUNTRY	3	4%
DISCO	3	4%
HIP-HOP	1	1%
JAZZ	4	5%
MÉTAL	1	1%
POP	0	0%
REGGAE	0	0%
ROCK	0	0%
Total	81	

RÉSULTATS DE LA REQUÊTE EXPLOITANT 4 RÈGLES D'ASSOCIATION

Règles 1 : '797;' → '517;'

Règles 2 : '323;' → '517;'

Règles 3 : '625;' → '517;'

Règles 4 : '410;' → '517;'

Genre	Document	Pourcentage
BLUES	11	14%
CLASSIQUE	56	72%
COUNTRY	2	3%
DISCO	3	4%
HIP-HOP	1	1%
JAZZ	4	5%
MÉTAL	1	1%
POP	0	0%
REGGAE	0	0%
ROCK	0	0%
Total	78	

RÉSULTATS DE LA REQUÊTE EXPLOITANT 5 RÈGLES D'ASSOCIATION

Règles 1 : '797;' → '517;'

Règles 2 : '323;' → '517;'

Règles 3 : '625;' → '517;'

Règles 4 : '410;' → '517;'

Règles 4 : '410;' → '625;'

Genre	Document	Pourcentage
BLUES	9	13%
CLASSIQUE	54	75%
COUNTRY	1	1%
DISCO	3	4%
HIP-HOP	1	1%
JAZZ	4	6%
MÉTAL	0	0%
POP	0	0%
REGGAE	0	0%
ROCK	0	0%
Total	72	

BIBLIOGRAPHIE

- [1] Alexander Schindler & Andreas Rauber (2012), Capturing the Temporal Domain in Echnonest Features for Improved Classification Effectiveness, In Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation, pp. 214-227
- [2] Aniruddh D. Patel (2003), A New Approach to the Cognitive Neuroscience of Melody, In The Cognitive Neuroscience of Music, Oxford University Press, pp 325-345.
- [3] Amihoud Amir, Yonatan Aumann, Ronen Feldman & Moshe Fresko (2005), Maximal Association Rules: A Tool for Mining Associations in Text, Journal of Intelligent Information System, 25:3, p. 333-345.
- [4] Asteris Zacharakis, Kostantinos Pastiadis, Georgios Papadelis, Joshua D. Reiss (2011), An invertigation of musical timbre: uncovering salient semantic descriptors and perceptual dimensions, In Proceeding of ISMIR 2011, Miami, U.S.A.
- [5] Avery Li-Chun Wang (2003), An Industrial-Strength Audio Algorithm, In Proceedings of ISMIR 2003, pp. 7-13.
- [6] A. C. North and D. J. Hargreaves (2008). The social and applied psychology of music. Oxford: Oxford University Press.
- [7] B. K. Baniya, J. Lee (2014), Importance of audio feature reduction in automatic music genre classification. Multimedia Tools and Applications, pp. 1-14.
- [8] B. L. Sturm (2012). A survey of evaluation in music genre recognition. In Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation, Springer International Publishing, pp. 29-66.
- [9] Ben Congleton, John Booker, Laurian C. Vega, Meg Kurdziolek, Lauren Shupp & Manuel A. Pérez-Quñones (2007), CA3: collaborative annotation of audio in academia, In Proceedings of ACM-SE 45.

- [10] Benoît Vaillant (2006), Mesurer la qualité des règles d'association : études formelles et expérimentales, Thèse présentée à l'École Nationale Supérieure des Télécommunications de Bretagne.
- [11] Benjamin Martin (2012), Analyse de structures répétitives dans les séquences musicales, Thèse de doctorat déposée à l'Université de Bordeaux 1.
- [12] Bernard Caillaud et Mireille Leriche (1999) Analyse Sonographique et aspect de la phonétique appliquée, Revue de l'EPI, pp.57-70.
- [13] Brian A. Whitman (2005), Learning the Meaning of Music, Ph.D thesis submitted to the Program in Media Arts and Sciences, MIT.
- [14] Carlos Pérez-Sancho, David Rizo and José M. Iñesta (2009), Genre classification using chords and stochastic language models, Connection Science, Vol. 21, No. 2, pp. 145 -159.
- [15] C. E. Shannon (1948), A Mathematical Theory of Communication, Bell System Technical Journal.
- [16] C.S.Kanimozhi Selvi & A.Tamilarasi (2009), An Automated Association Rule Mining Technique With Cumulative Support Thresholds, Int. J. Open Problems in Comput. Math, Vol. 2, No. 3
- [17] C.S.Kanimozhi Selvi & A.Tamilarasi (2011), Mining Rare Itemset with Automated Support Thresholds, Journal of Computer Science, Vol 7, No. 3, p 394-399.
- [18] Charles Bouveyron (2006): Modélisation et classification des données de grandes dimensions, Application à l'analyse d'images. Thèse de Doctorat déposée à l'Université Joseph Fourier, France.
- [19] Chris Adamson & Kevin Avila (2012), Learning Core Audio : A Hands-On Guide to Audio Programming for Mac and iOS, Addison-Wesley Professional.
- [20] Cory McKay & Ichiro Fujinaga (2008), Combining features extracted from audio, symbolic and cultural sources, In Proceedings of ISMIR 2008.
- [21] Cory McKay & Ichiro Fujinaga (2004), Automatic genre classification using large high-level musical feature sets, In Proceedings of ISMIR 2004.

- [22] Cory McKay (2010) Automatic music classification with jMIR. Ph.D. Thesis. McGill University, Canada.
- [23] C. McKay, I. Fujinaga, & P. Depalle (2005). jAudio: A feature extraction library. In Proceedings of the International Conference on Music Information Retrieval.
- [24] Daniel J. Levitin (2006), This is your brain on music, Plume, ISBN 978-0-452-28852-2.
- [25] Daniel J. Levitin et Ursula Bellugi (2006), Rhythm, Timbre, and Hyperacusis in Williams-Beuren Syndrome, Chapitre 16 dans Williams-Beuren Syndrome: Research and Clinical Perspectives, Johns Hopkins University Press, 343-358.
- [26] Daniel J. Levitin & Anna K. Tirovolas (2010), Music Cognition and Perception, In Sage Encyclopedia of Perception, Sage Publication, pp. 599-606.
- [27] Daniel McEnnis, Cory McKay, Ichiro Fujinaga & Philippe Depalle (2005), jAudio : A Feature Extraction Library, In Proceeding of ISMIR 2005.
- [28] Darrel Conklin (2009), Melody classification using patterns. In Second International Workshop on Machine Learning and Music (pp. 37-41).
- [29] David A. Levitt (1985), A Representation for Musical dialects, Thesis, Massachusetts Institute of Technology.
- [30] Elias Pampalk, Simon Dixon & Gerhard Widmer (2003), Exploring Music Collection by Browsing Different Views, In Proceedings of ISMIR 2003, Baltimore, U.S.A.
- [31] Erling Wold, Thom Blum, Douglas Keislar & James Wheaton (1996), Content-Based Classification, and Retrieval of Audio, Multimedia IEEE, Volume 3, Issue 3, p. 27-36
- [32] E. Rosch & C. B. Mervis (1975), Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, pp. 573-605.
- [33] Fu-Hai Frank Wu, Tsung-Chi Lee, Jyh-Shing Roger Jang, Kaichun K. Chang, Chun Hung Lu & Wen Nan Wang (2011), A Two-Fold Dynamic

programming approach to beat tracking for audio music with time-varying tempo, In Proceeding of ISMIR 2011, Miami, U.S.A.

- [34] Geoffrey E. Hinton, Simon Osindero & Yee-Whye Teh, (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.
- [35] Geoffrey E. Hinton (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE*, 29(6), 82-97.
- [36] George Tzanetakis (2002) *Manipulation, Analysis and Retrieval Systems for Audio Signals*. PhD thesis, Princeton University.
- [37] George Tzanetakis & Perry Cook (2002), Musical Genre Classification of Audio Signals, In *IEEE Transactions on Speech and Audio Processing*, Vol. 10, No. 5, pp. 293-302
- [38] George Tzanetakis and Perry Cook. 2000. MARSYAS: A framework for audio analysis. *Organized Sound* 4 (3): 169–75.
- [39] George Tzanetakis, Georg Essl & Perry Cook (2002), Automatic musical genre classification of audio signals, In *IEEE Transactions on speech and audio processing*.
- [40] Graham E. Poliner, Daniel D. P. Ellis, Andreas F. Ehmann, E. Gómez, S. Streich & B. Ong (2007). Melody transcription from music audio: Approaches and evaluation. *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 15, No 4, pp. 1247-1256.
- [41] Greg Linden, Brent Smith & Jeremy York (2003), Amazon.com Recommendations: Item-to-Item Collaborative Filtering, *Internet Computing IEEE*, Volume 7, Issue 1, pp. 76-80
- [42] Geoffroy Peeters (2004), A large set of audio features for sound description (similarity and classification) in the CUIDADO project, *IRCAM Report Paper*.
- [43] Host Eidenberger (2012) *Handbook of Multimedia Information Retrieval*, atpress, Vienna, ISBN 978-3848222834.

- [44] Hrishikesh Deshpande, Rohit Singh & Unjung Nam (2001), Classification of Music Signals in The Visual Domain, In Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01), Limerick, Ireland
- [45] Ismaïl Biskri, Hassane Hilali & Louis Rompré (2010), Extraction de relations d'association maximales dans les textes, Actes du JADT 2010, p. 173-182.
- [46] Ismaïl Biskri, Jean-Guy Meunier (2002), SATIM : Système d'Analyse et de Traitement de l'Information Multidimensionnelle, Actes du colloque JADT 2002, St-Malo, France.
- [47] Ismaïl Biskri, Louis Rompré, Lamri Laouamer & François Meunier (2006), Classification de document multimédias : vers une approche générale. Acte de la conférence JADT 2006. Besançon, France.
- [48] Ismaïl Biskri & Louis Rompré (2012), Using association rules for query reformulation. In Next Generation Search Engines: Advanced Models for Information Retrieval, Eds Christophe Jouis, Ismaïl Biskri, Jean-Gabriel Ganascia & Magali Roux. IGI GLOBAL Press.
- [49] Ismaïl Biskri, Louis Rompré, Steve Descoteaux, Abdelghani Achouri, Boucif Amar Bensaber, «Extraction of Strong Associations in Classes of Similarities», In proceedings of the 11th International Conference on Machine Learning and Applications IEEE/ICMLA 2012, December 12-15, Boca Raton, Florida, USA
- [50] James A. Anderson (1995), An Introduction to Neural Network, MIT Press, ISBN 0-262-01144-1.
- [51] Jayme Garcia Arnal Barbedo & Amauri Lopes (2007) Automatic Genre Classification of Music Signals, In Journal on Advances in Signal Processing.
- [52] Jean-Julien Aucouturier & François Pachet (2003), Representing Musical Genre : A State of Art, In Journal of New Music Research, Vol. 32, No. 1 pp. 83-93.
- [53] Jean-Julien Aucouturier & François Pachet (2003), Representing Musical Genre : A State of Art, In Journal of New Music Research, Vol. 32, No. 1 pp. 83-93.

- [54] Jean-Guy Meunier, Ismaïl Biskri & Dominic Forest (2005), A model for computer analysis and reading of text: the SATIM approach, In Journal of computer text processing Text Technology, Vol. 14, No. 2, pp. 123 – 151.
- [55] J. S, Downie, A. Ehmann, & D. Tcheng (2005), Music-to-knowledge (M2K): A prototyping and evaluation environment for music information retrieval research. Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval. 676.
- [56] Julien Piquier (2004), Indexation sonore : recherche de composantes primaires pour une structuration audiovisuelle, Thèse de doctorat déposé à l'Université Toulouse III.
- [57] Jiawei Han, Jian Pei, Yiwen Yin, Runying Mao (2004), Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach, Data Mining and Knowledge Discovery, 8, pp 53-87.
- [58] John Cook (1994), Agent Reflection in an Intelligent Learning Environment Architecture for Composition, In Music Education: An Artificial Intelligence Approach, Springer Verlag.
- [59] Kris West & Paul Lamere (2007), A Model-Based Approach to Constructing Music Similarity Functions, In Journal on Advances in Signal Processing.
- [60] Lie Lu, Hao Jiang & Hongjiang Zhang (2001), A Robust Audio Classification and Segmentation Method, In Proceedings of ACM international conference on Multimedia, pp 203 – 211
- [61] Lisianne Hoch, Barbara Tillmann et Bénédite Poulin-Charronnat (2008), Musique, syntaxe et sémantique : des ressources d'intégration structurale et temporelle partagées ? Revue de Neuropsychologie 2008, Vol. 18, No 1-2, 33-59.
- [62] Ling Chen, Phillip Wright & Wolfgang Nejdl (2009), Improving Music Genre Classification Using Collaborative Tagging Data, In Proceedings of the Second ACM International Conference on Web Search and Data Mining.

- [63] Liquiang Geng & Howard J. Hamilton (2006), Intertingness Measures for Data Mining: A Survey, In Journal ACM Computing Surveys, Vol. 38, Issue 3.
- [64] Louis Rompré, Ismaïl Biskri & François Meunier (2008), Text classification : A preferred tool for audio file classification. Acte de la conférence AICCSA 2008 IEEE/ACS. Doha, Qatar.
- [65] Louis Rompré, Ismaïl Biskri & François Meunier (2008), La classification textuelle : un outil privilégié pour la classification de documents audio, Actes de la conférence JADT, Lyon, France.
- [66] Marc Parizau (2006), Réseaux de neurones, Université Laval.
- [67] Mark F. Bear, Barry W. Connors & Michael A. Paradiso (2007), Neurosciences: à la découverte du cerveau (3e édition), Les Éditions Pradel.
- [68] Mohammed J. Zaki, Wagner Meira, Jr., (2014) Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, May 2014. ISBN: 9780521766333
- [69] Murat Kunt (2000), Reconnaissance des formes et analyse de scènes, volume 3 : Traitement de l'information, Presses Polytechniques et Universitaires Romandes, ISBN 978-28807-7438-0.
- [70] Nicholas Pasquier (2000), Data Mining : Algorithmes d'extraction et de réduction des règles d'association dans les bases de données, Thèse de Doctorat déposée à l'université de Clermont-Fernand II, France.
- [71] Nicolas Ruwet (1966) Méthodes d'analyse en musicologie. Revue belge de Musicologie/Belgisch Tijdschrift voor Muziekwetenschap, pp. 65-90.
- [72] Olivier Lartillot et Petri Toiviainen (2007), A Matlab toolbox for musical feature extraction from audio. In Proceeding of the 10th International conference on Digital Audio Effects, Bordeaux, France.
- [73] Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2006), Introduction to Data Mining, Addison-Wesley, ISBN 978-03213-2136-7

- [74] Pang-Ning Tan, Vipin Kumar, Jaideep Srivastava (2002), Selecting the Right Interestingness Measure for Association Patterns, In Proceedings of ACM SIGKDD 02, pp. 32-41.
- [75] Perfecto Herrera, Alexandre Yeterian & Fabien Gouyon (2002), Automatic Classification of Drum Sounds: A Comparison of Feature Selection Methods and Classification Techniques, In Lecture Notes in Computer Science, Vol. 2445, pp. 69-80
- [76] Perter Grosche, Meinard Müller & Joan Serrà (2012) Audio Content-Based Music Retrieval, In Multimodal Music Processing, vol. 3, pp. 157-174
- [77] Peter J. Rentfrow, Lewis R. Goldberg & Daniel J. Levitin (2011), The Structure of Musical Preferences: A Five-Factor Model, Journal of Personality and Social Psychology, Vol. 100, No. 6, 1139–1157
- [78] P. Lenca, P. Meyer, B. Vaillant, et S. Lallich (2008). On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid. European Journal of Operational Research 184, p 610–626.
- [79] Quoc V. Le et al. (2012). Building High-level Features Using Large Scale Unsupervised Learning.
- [80] Rakesh Agrawal, Tomasz Imielinski & Arun Swami (1993), Mining association rules between sets of items in large databases. In proceedings of the ACM SIGMOD Conference on Management of Data : p. 207-216.
- [81] Rakesh Agrawal & Ramakrishnan Srikant (1994), Fast Algorithms for Mining Association Rules, In Poceedings of the 20th International Conference on Very Large Database, p. 487-499.
- [82] Rémi Foucard, Jean-Louis Durrieu, M. Lagrange et Gaël Richard (2010), Multimodal similarity between musical streams for cover version detection, In Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, ICASSP 2010, Dallas, U.S.A.
- [83] Richard S. Sutton & Andrew G. Barto (1998), Reinforcement Learning : An Introduction, MIT Press, ISBN 0-262-19398-1.
- [84] Roger S. Pressman (2001), Software Engineering : A Practitioner's Approach, (5e Edition) McGraw-Hill, ISBN 0-07-249668-9.

- [85] Ronen Feldman, Yonatan Aumann, Amihod Amir, Amir Zilberstein & Willi Kloesgen (1997), Maximal Association Rules: a New Tool for Mining for Keyword co-occurrences in Document Collections, In proceedings of the KDD-97, Newport Beach, USA.
- [86] Rory A. Lewis & Alicja Wieczorkowska (2007), Categorization of Musical Instrument Sounds Based on Numerical Parameters, In ICCS 2007, pp 87-93
- [87] Sally P. Springer & Georg Deutsch (2001), Left Brain, Right Brain. Perspectives from Cognitive Neuroscience, Fifth edition, W.H. Freeman and Company/Worth Publishers, ISBN 978-0-716-73111-5.
- [88] Scott Beveridge & Don Know (2012), A Feature Survey for Emotion Classification of Western Popolar Music, In 9th International Symposium on Computer Music Modelling and Retrivial (CMMR 2012), London, UK.
- [89] Siddharth Sigtia & Simon Dixon (2014), Improved Music Learning With Deep Neural Network, n Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on (pp. 6959-6963).
- [90] Simon Dixon, Elias Pampalk & Gerhard Widmer (2003), Classification of Dance Music by Periodicity Patterns, In Proceedings of ISMIR 2003, Baltimore, U.S.A.
- [91] Simon Haykin (1994), Neural Networks: A Comprehensive Foundation, Macmillan College Publishing Company, ISBN 0-02-352761-7.
- [92] Simon Holland (2000), Artificial Intelligence in music education: a critical review, In Readings in Music and Artificial Intelligence, Contemporary Music Studies, Volume 20, ISBN 90-5755-094-6, pages 239-274.
- [93] Shi-Huang Chen & Shih-Hao Chen (2009) Content-based Music Genre Classification Using Timbral Feature Vectors and Support Vector Machine, In Proceedings of ICIS 2009.
- [94] Stanley Smith Stevens, John Volkman and Edwin Newman (1937) A scale for the measurement of psychological magnitude pitch. Journal of the Acoustical Society of America, 8: 185-190.

- [95] Sylvie Guillaume, Ali Khenchaf (2000), Règles ordinales : une généralisation des règles d'association. Application, Acte des Journées sur la fouille dans les données par la méthode d'analyse statistique implicative, Edition des Actes coordonnée par Régis Gras et Marc Bailleul.
- [96] S. Bray, and G. Tzanetakis (2005), Distributed audio feature extraction for music. Proceedings of the International Conference on Music Information Retrieval. 434–7.
- [97] Takahiro Tada, Tomomasa Nagashima & Yoshifumi Okada (2010), Rule-based classification for audio data based on closed itemset mining, In Proceedings of IMECS 2010, Hong Kong.
- [98] Tao Li, Mitsunori Ogihara & Qi Li (2003), A comparative study on content-based music genre classification, In proceedings of the 26th annual international ACM SIGIR conference on Research and development in information Retrieval. pp. 282-289.
- [99] The Nielsen Soundscan Company (2011), Nielsen Soundscan 2011 Music Industry Report, <http://nielsen.com/global/en.html>
- [100] Thierry Bertin-Mahieux, Douglas Eck & Michael Mandel (2010), Automatic Tagging of Audio: The State-of-the-Art, In Wenwu Wang, editor, Machine Audition: Principles, Algorithms and Systems. IGI Publishing.
- [101] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman & Paul Lamere (2011), The Million Song Dataset, In Proceedings of ISMIR 2011, Miami, USA.
- [102] Tom Arjannikov & John Z. Zhang (2014), An Association-Based Approach to Genre Classification in Music. In Proceedings of ISMIR 2014, Taipei, Taiwan.
- [103] Thomas J. Cavicchi (2000). Digital Signal Processing. John Wiley & Sons, Inc.
- [104] Tristan Jehan (2005), Creating Music by Listening, Ph.D thesis submitted to the Program in Media Arts and Sciences, MIT.
- [105] Vikramjit Mitra & Chia J. Wang (2007), A Neural Network based Audio Content Classification, TENCON 2007 IEEE Region 10 Conference.

- [106] V. Vapnik (1995), The Nature of Statistical Learning Theory. Springer-Verlag, New York.
- [107] Vinjamuri Swathi, M.Sunitha Reddy (2014), Music Recommendation System Using Association Rules, In International Journal of Technology enhancements and emerging engineering research, Vol. 2, Issue 7.
- [108] W. K. Estes (1994), Classification and Cognition, Oxford University Press, ISBN 0-19-510974-0.
- [109] Xavier Glorot, Antoine Bordes & Yoshua Bengio (2011), Deep sparse rectifier neural networks. In International Conference on Artificial Intelligence and Statistics (pp. 315-323).
- [110] Yann LeCun, Yoshua Bengio & Geoffrey E. Hinton, (2015). Deep learning. Nature, 521(7553), 436-444.
- [111] Yoshua Bengio, Pascal Lamblin, Dan Popovici and Hugo Larochelle (2007), Greedy Layer-Wise Training of Deep Networks, in J. Platt et al. (Eds), Advances in Neural Information Processing Systems 19 (NIPS 2006), pp. 153-160, MIT Press.
- [112] Y. Le Bras, P. Meyer, P. Lenca et S. Lallich (2010), Mesure de la robustesse de règles d'association, Actes du QDC 2010, Hammamet, Tunisie.
- [113] Zeynab Liraki, Ali Harounabadi & Javad Mirabedini (2015), Predicting the Users'Navigation Patterns in Web, using Weighted Association Rules and Users' Navigation Information, In International Journal of Computer Application (0975-8887), Volume 110 – No. 12.